

Change detection in INAR(p) processes against various alternative hypotheses

GYULA PAP AND TAMÁS T. SZABÓ

Bolyai Institute, University of Szeged, Szeged, Hungary

Address of correspondence: Tamás T. Szabó, Bolyai Intézet, Aradi vértanúk tere 1, Szeged, H-6720,

Hungary; E-mail: tszabo@math.u-szeged.hu

Abstract

Change in the coefficients or in the mean of the innovation of an INAR(p) process is a sign of disturbance that is important to detect. The proposed methods can test for change in any one of these quantities separately, or in any collection of them. They make both one-sided and two-sided tests possible, furthermore, they can be used to test against the ‘epidemic’ alternative. The tests are based on a CUSUM process using CLS estimators of the parameters. Under the one-sided and two-sided alternatives consistency of the tests is proved and the properties of the change-point estimator are also explored.

Keywords: Change-point detection; stable INAR(p) process

2010 AMS subject classifications: primary 62M02; secondary 60J80, 60F17

1 Introduction

A time-inhomogeneous INAR(p) process is a sequence $(X_k)_{k \geq -p+1}$ given by

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{1,k,j} + \cdots + \sum_{j=1}^{X_{k-p}} \xi_{p,k,j} + \varepsilon_k, \quad k \in \mathbb{N}, \quad (1.1)$$

where $\{\varepsilon_k : k \in \mathbb{N}\}$ is a sequence of independent non-negative integer-valued random variables, for each $k \in \mathbb{N}$ and $i \in \{1, \dots, p\}$ the sequence $\{\xi_{i,k,j} : j \in \mathbb{N}\}$ is a sequence of i.i.d. Bernoulli random variables with mean $\alpha_{i,k}$ such that these sequences are mutually independent and independent of the sequence $\{\varepsilon_k : k \in \mathbb{N}\}$, and X_0, \dots, X_{-p+1} are non-negative integer-valued random variables independent of the sequences $\{\xi_{i,k,j} : j \in \mathbb{N}\}$, $k \in \mathbb{N}$, $i \in \{1, \dots, p\}$, and $\{\varepsilon_k : k \in \mathbb{N}\}$. The numbers $\alpha_{i,k}$ are called coefficients, and we will refer to $\varepsilon_1, \varepsilon_2, \dots$ as the innovations. Time-homogeneous INAR(p) processes have a number of applications, which are summarized, e.g., in Barczy et al. (2011).

The reason that we initially define our process as time-inhomogeneous is that we would like to test for a change in the parameters, therefore we have to allow them to vary over time. In the proofs, however, a majority of the results will be based upon the properties of time-homogeneous INAR(p) processes.

In applications the time series underlying an observation is usually supposed to be homogeneous, and this can lead to false conclusions if the parameters have changed during the time of the observation. Therefore testing for a change in the parameters has been an important question. The monograph of Csörgő and Horváth (1997) gives an excellent overview of the subject for many times series models, and in their own context, their main results are stronger than those of this paper—e.g., in place of Theorem 4.2 they usually prove convergence in distribution under a variety of conditions. Their methods, however, cannot be directly used to handle $\text{INAR}(p)$ processes—they routinely involve a Taylor expansion of the likelihood function, which we cannot do because we are not assuming that the innovation distribution comes from a parametric family. Furthermore, the autoregressive processes investigated in Csörgő and Horváth (1997) have the usual type of autoregression—i.e., the next element of the time series is a measurable function of the previous elements, plus an independent innovation. In our case, however, the next element depends stochastically on the previous elements as well. Therefore, it is more fruitful to consider our model as a special multitype branching process. This will be our approach as well, but we remark that much of the motivation for the paper is due to Csörgő and Horváth (1997) and Gombay (2008). We note that our Theorem 3.1 and Theorem 4.1 are weaker than the equivalent versions of Theorems 1 and 2 in Gombay (2008)—this is due to our model being more complicated. In our continued research we plan to improve our results in this direction. Similarly, in place of Theorem 4.2 we hope to show that the difference $\tau_n - \tau$ converges in distribution, as it was shown in Csörgő and Horváth (1997) for the simpler models. Finally, we remark that it is indeed proper to take motivation from these sources because our process resembles an $\text{AR}(p)$ process in its covariance structure.

Change detection methods for $\text{INAR}(p)$ processes in general (i.e., with no prespecified innovation distribution) have only been proposed in a few papers—we refer to Kang and Lee (2009) especially, where the authors give a test statistics similar to ours for a more general model. However, no result is available under the alternative hypothesis and the asymptotics of the change-point estimator are not given—we will give results to these questions which strengthen the theoretical foundations of the method considerably.

The test has already been published in T. Szabó (2011b). However, the referred paper only states the main result but does not provide the proof. Also, it contains no results under the alternative hypothesis either, which is the main result of this paper. In the appendix we will also provide the missing (although rather standard) proof in T. Szabó (2011b).

Now we proceed with the formulation of the statistical problem. We assume $\mu_k := \mathbb{E}(\varepsilon_k) < \infty$ and $0 < \sigma_k^2 := \text{Var}(\varepsilon_k) < \infty$.

Write the parameter vectors as

$$\begin{bmatrix} \alpha_{1,k} \\ \vdots \\ \alpha_{p,k} \\ \mu_k \end{bmatrix} =: \boldsymbol{\theta}_k,$$

and let us choose a subset PV of $\{1, 2, \dots, p+1\}$ such that

$$PV = \{i_1, i_2, \dots, i_\ell\} \text{ for some } \ell > 0, \quad i_1 < i_2 < \dots < i_\ell. \quad (1.2)$$

Also, we can write

$$NV := PV^c = \{j_1, j_2, \dots, j_{p+1-\ell}\}, \quad j_1 < j_2 < \dots < j_{p+1-\ell},$$

where H^c denotes the complement of a set. Let us now define

$$\boldsymbol{\varphi}_k := (\theta_k^{(i_1)}, \theta_k^{(i_2)}, \dots, \theta_k^{(i_\ell)})^\top, \quad \boldsymbol{\eta}_k := (\theta_k^{(j_1)}, \theta_k^{(j_2)}, \dots, \theta_k^{(j_{p+1-\ell})})^\top.$$

The vector $\boldsymbol{\varphi}_k$ is the parameter vector of interest and $\boldsymbol{\eta}_k$ is the 'nuisance' parameter vector. For a fixed number of observations n we want to test the null hypothesis

$$H_0 : \quad \varepsilon_1, \dots, \varepsilon_n \text{ are identically distributed and } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_n$$

against the alternative

$$\begin{aligned} H_A : \quad & \text{there is an integer } \tau \in \{1, \dots, n-1\} \text{ such that} \\ & \boldsymbol{\varphi}_1 = \dots = \boldsymbol{\varphi}_\tau \neq \boldsymbol{\varphi}_{\tau+1} = \dots = \boldsymbol{\varphi}_n \text{ but } \boldsymbol{\eta}_1 = \dots = \boldsymbol{\eta}_n, \\ & \varepsilon_1, \dots, \varepsilon_\tau \text{ are identically distributed,} \\ & \text{and } \varepsilon_{\tau+1}, \dots, \varepsilon_n \text{ are identically distributed.} \end{aligned}$$

The main difficulty is that the change is not in a location or scaling parameter as in most of the change detection models, and standard techniques based on likelihood are not applicable.

Under the null hypothesis H_0 we have

$$\begin{bmatrix} \alpha_{1,1} \\ \vdots \\ \alpha_{p,1} \\ \mu_1 \end{bmatrix} = \dots = \begin{bmatrix} \alpha_{1,n} \\ \vdots \\ \alpha_{p,n} \\ \mu_n \end{bmatrix} =: \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\alpha} \\ \mu \end{bmatrix} =: \boldsymbol{\theta}, \quad \sigma_1^2 = \dots = \sigma_n^2 =: \sigma^2,$$

and the process (1.1) is referred to as stable, unstable or explosive whenever $\alpha_1 + \dots + \alpha_p < 1$, $\alpha_1 + \dots + \alpha_p = 1$ or $\alpha_1 + \dots + \alpha_p > 1$, respectively. Basic differences between the three types are summarized in Barczy et al. (2011). We will study only the stable case $\alpha_1 + \dots + \alpha_p < 1$ with nonvanishing innovation (i.e. with $\mu > 0$), when the Markov chain $(\mathbf{X}_k)_{k \geq 0}$ given by

$$\mathbf{X}_k := \begin{bmatrix} X_k \\ \vdots \\ X_{k-p+1} \end{bmatrix}, \quad k \geq 0,$$

has a nondegenerate, unique stationary distribution (see e.g. Quine (1970)). Under the alternative hypothesis we will require that the process be stable both before and after the change. This assumption is central to our proofs, because if the process is not stable, then it is not ergodic either and we cannot use any of the methods outlined below. For unstable and explosive processes completely different tools should be developed.

The structure of the paper is the following. In the second section we propose a CUSUM-like test process similar to that of Kang and Lee (2009), in the third section we compute its limiting distribution and, based on the process, we introduce both one-sided and two-sided tests. For simulations on the power of these tests we refer the reader to T. Szabó (2011b). Our main results under the alternative hypothesis are stated in Section 4 and a real data illustration is given in Section 5. Appendix A contains detailed proofs under the null hypothesis, while Appendix B contains the proofs under the alternative hypothesis. The paper is concluded by Appendix C, which contains some necessary technical lemmas and calculations. Unless otherwise noted, we understand convergence as $n \rightarrow \infty$. We will denote the set of positive integers by \mathbb{N} and the set of nonnegative integers by \mathbb{N}_0 .

2 Parameter estimation and the construction of our test process

2.1 Parameter estimates

The first step in parameter estimation is introducing the sequence of martingale differences $(M_k)_{k \in \mathbb{N}}$:

$$M_k = X_k - E(X_k | \mathcal{F}_{k-1}) = X_k - \boldsymbol{\alpha}^\top \mathbf{X}_{k-1} - \mu, \quad k \in \mathbb{N}, \quad (2.1)$$

where $(\mathcal{F}_n)_{n \in \mathbb{N}}$ is the natural filtration. The conditional least squares estimators of the parameters, first introduced by Klimko and Nelson (1978), can be calculated by minimizing the sum of squares

$$R_n(\alpha_1, \dots, \alpha_p, \mu) := \frac{1}{2} \sum_{k=1}^n M_k^2 = \frac{1}{2} \sum_{k=1}^n (X_k - \boldsymbol{\alpha}^\top \mathbf{X}_{k-1} - \mu)^2$$

with respect to $\alpha_1, \dots, \alpha_p, \mu$. With a reasoning completely analogous to that of 3.1 Lemma and 3.1 Proposition in Barczy, Ispány, and Pap (2012) we can show that R_n has a unique minimum given by (2.2) whenever \mathbf{Q}_n is invertible. Now, we will show in C.1 that this is true with an asymptotic probability of 1, therefore the parameter estimates exist and are unique with an asymptotic probability of 1. As all our results are asymptotic, this will be sufficient for the purposes of the paper.

$$\hat{\boldsymbol{\theta}}_n := \begin{bmatrix} \hat{\boldsymbol{\alpha}}_n \\ \hat{\mu}_n \end{bmatrix} := \mathbf{Q}_n^{-1} \sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}, \quad \mathbf{Q}_n := \sum_{k=1}^n \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top. \quad (2.2)$$

We note that this estimator is strongly consistent under the null hypothesis (Du & Li, 1991), and the estimation procedure itself supposes that the null hypothesis is valid; however, the calculations can be carried out under the alternative hypothesis as well. Under the alternative hypothesis the weak limit of $\hat{\boldsymbol{\theta}}_n$ is given in Lemma B.3. Replacing the parameters by their estimates in M_k we obtain $\widehat{M}_k^{(n)}$, i.e.,

$$\widehat{M}_k^{(n)} := X_k - \hat{\boldsymbol{\theta}}_n^\top \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}. \quad (2.3)$$

Although not a parameter in which we are looking for change, the estimate of the variance of the innovation σ^2 will also appear in our test process, therefore we have to provide an estimator for it. To do this, we introduce

$$N_k = M_k^2 - E(M_k^2 | \mathcal{F}_{k-1}) = M_k^2 - \alpha_1(1 - \alpha_1)X_{k-1} - \dots - \alpha_p(1 - \alpha_p)X_{k-p} - \sigma^2, \quad k \geq 0.$$

Minimizing $\sum_{k=1}^n N_k^2$ with respect to σ^2 we obtain the conditional least squares estimate

$$\bar{\sigma}_n^2 = -\frac{1}{n} \sum_{k=1}^n (M_k^2 - \alpha_1(1 - \alpha_1)X_{k-1} - \cdots - \alpha_p(1 - \alpha_p)X_{k-p}). \quad (2.4)$$

However, in this estimate the true parameters are still present. The estimate that we will use is given by replacing the α coefficients and μ both in the formula and in M_k^2 by their estimates:

$$\hat{\sigma}_n^2 = -\frac{1}{n} \sum_{k=1}^n \left(\left(\widehat{M}_k^{(n)} \right)^2 - \hat{\alpha}_1^{(n)} \left(1 - \hat{\alpha}_1^{(n)} \right) X_{k-1} - \cdots - \hat{\alpha}_p^{(n)} \left(1 - \hat{\alpha}_p^{(n)} \right) X_{k-p} \right). \quad (2.5)$$

2.2 Construction of the test process

We use a formal analogy between the INAR(p) process and the well-known AR(p) process (Venkataraman, 1982) to obtain analogues of score vector and information quantities as in Gombay (2008). We briefly recall the motivation of the test process as given in T. Szabó (2011b). We briefly recall the motivation of the test process as given in T. Szabó (2011b). Due to the martingale central limit theorem,

$$\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{\lceil nt \rceil} M_k \right)_{t \in [0,1]} \xrightarrow{\mathcal{D}} (\sqrt{c} \mathcal{W}_t)_{t \in [0,1]},$$

where c is a constant depending on $\boldsymbol{\theta}$ and σ^2 , and $(\mathcal{W}_t)_{0 \leq t \leq 1}$ is a standard Brownian motion. Therefore, by a rough approximation

$$(M_1, \dots, M_n) \sim N(0, cE_n),$$

where E_n is the $n \times n$ identity matrix. The approximate likelihood function is

$$\frac{1}{(2\pi c)^{n/2}} \exp \left\{ -\frac{1}{2c^2} \sum_{k=1}^n M_k^2 \right\}.$$

We will take the derivative of the log-likelihood function and work with that quantity. The first term will be regarded as constant. This is a simplification because c actually depends on the parameters but taking this into account leads to calculations that are difficult to handle. Also, we will not take into account the constant factor before the sum of the M_k but will rather work with the analogue of the information matrix. Therefore, we consider the following analogue of the loglikelihood function:

$$R_n(\alpha_1, \dots, \alpha_p, \mu) = -2^{-1} \sum_{k=1}^n M_k^2.$$

Note that this is the sum that we had to minimize for CLS estimations. The role of the score vector will be played by

$$-\nabla R_k(\hat{\boldsymbol{\theta}}_n) = \sum_{j=1}^k \widehat{M}_j^{(n)} \begin{bmatrix} \mathbf{X}_{j-1} \\ 1 \end{bmatrix}.$$

The information matrix \mathbf{I}_n is defined by

$$\mathbf{I}_n := \sum_{k=1}^n \mathbb{E}[\{\nabla R_k(\boldsymbol{\theta}) - \nabla R_{k-1}(\boldsymbol{\theta})\} \{\nabla R_k(\boldsymbol{\theta}) - \nabla R_{k-1}(\boldsymbol{\theta})\}^\top | \mathcal{F}_{k-1}] = \sum_{k=1}^n (\boldsymbol{\alpha}_2^\top \mathbf{X}_{k-1} + \sigma^2) \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top,$$

with

$$\boldsymbol{\alpha}_2 := [\alpha_1(1 - \alpha_1), \dots, \alpha_p(1 - \alpha_p)]^\top. \quad (2.6)$$

The estimate $\widehat{\mathbf{I}}_n$ is defined by replacing in \mathbf{I}_n the variance σ^2 and all the parameters in $\boldsymbol{\theta}$ by their CLS estimates. This leads to the $p + 1$ -dimensional test process $(\widehat{\mathcal{M}}_n(t))_{0 \leq t \leq 1}$ given by

$$\widehat{\mathcal{M}}_n(t) := \widehat{\mathbf{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{M}_k^{(n)} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}. \quad (2.7)$$

Note that the process $(\widehat{\mathcal{M}}_n(t))_{0 \leq t \leq 1}$ can also be written in the CUSUM form

$$\begin{aligned} \widehat{\mathcal{M}}_n(t) &= \widehat{\mathbf{I}}_n^{-1/2} \left(\sum_{k=1}^{\lfloor nt \rfloor} X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} - \mathbf{Q}_{\lfloor nt \rfloor} \mathbf{Q}_n^{-1} \sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right) \\ &= \widehat{\mathbf{I}}_n^{-1/2} \mathbf{Q}_{\lfloor nt \rfloor} \left(\begin{bmatrix} \widehat{\boldsymbol{\alpha}}_{\lfloor nt \rfloor} \\ \widehat{\mu}_{\lfloor nt \rfloor} \end{bmatrix} - \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_n \\ \widehat{\mu}_n \end{bmatrix} \right) = \widehat{\mathbf{I}}_n^{-1/2} \mathbf{Q}_{\lfloor nt \rfloor} (\widehat{\boldsymbol{\theta}}_{\lfloor nt \rfloor} - \widehat{\boldsymbol{\theta}}_n). \end{aligned}$$

This is close to the CUSUM processes used by Kang and Lee (2009). In this paper the authors investigated a change in random coefficient INAR(p) processes based on both maximum-likelihood and conditional least squares estimators. A general convergence theorem was proved for a test process similar to ours based on a broader class of estimators. However, the authors did not state any results under the alternative hypothesis. Our Theorem 3.1 is analogous to Theorem 1 of Kang and Lee (2009). However, while their model is more general than ours, our Theorem 3.1 is not a consequence of their Theorem 1, even though the applied methods are largely identical.

3 Testing procedures

Definition. A time-homogeneous INAR(p) process $(X_k)_{k \geq -p+1}$ is said to satisfy condition \mathbf{C}_0 , if $E(X_0^6) < \infty$, \dots , $E(X_{-p+1}^6) < \infty$, $E(\varepsilon_1^6) < \infty$, $\alpha_1 + \dots + \alpha_p < 1$, $\mu > 0$ all hold for it, and if, furthermore, $\alpha_1 + \dots + \alpha_p > 0$ or $\sigma^2 > 0$.

An INAR(p) process $(X_k)_{k \geq -p+1}$ which satisfies H_A is said to satisfy condition \mathbf{C}_A , if $(X_k)_{-p+1 \leq k \leq \tau}$ and $(X_k)_{k \geq \tau+1}$ satisfy condition \mathbf{C}_0 .

Under the null hypothesis we have the following result, which allows the construction of various test statistics.

Theorem 3.1. *If $(X_k)_{k \geq -p+1}$ satisfies H_0 and condition \mathbf{C}_0 , then*

$$\widehat{\mathcal{M}}_n \xrightarrow{\mathcal{D}} \mathcal{B} \quad \text{as } n \rightarrow \infty,$$

where $(\mathcal{B}(t))_{0 \leq t \leq 1}$ is a $(p+1)$ -dimensional standard Brownian bridge, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution in the Skorokhod space $D([0, 1])$.

By the continuous mapping theorem we obtain the following corollary.

Corollary 3.2. *Under the assumptions of Theorem 3.1 we have*

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t), \quad (3.1)$$

$$\inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \xrightarrow{\mathcal{D}} \inf_{0 \leq t \leq 1} B(t), \quad (3.2)$$

$$\sup_{0 \leq t \leq 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} |B(t)|, \quad (3.3)$$

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) - \inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t) \quad (3.4)$$

as $n \rightarrow \infty$, where $(\widehat{\mathcal{M}}_n^{(i)}(t))_{0 \leq t \leq 1}$, $i = 1, \dots, p+1$, denotes the components of $(\widehat{\mathcal{M}}_n(t))_{0 \leq t \leq 1}$, and $(B(t))_{0 \leq t \leq 1}$ is a Brownian bridge.

Theorem 3.1 is proved in Appendix B.

Since $(\mathcal{B}(t))_{0 \leq t \leq 1}$ in Theorem 3.1 has independent components, we need to define the tests component-wise only. For simultaneous test-for-change in d parameters, to have an overall level of significance α , we use $\alpha^* := 1 - (1 - \alpha)^{1/d}$ for each component. We can test for change in a single component, $\theta^{(i)}$, $i \in PV$ (with $\theta^{(i)} = \alpha_i$ for $i = 1, 2, \dots, p$ and $\theta^{(p+1)} = \mu$, according to the definition of $\boldsymbol{\theta}$) in the following way:

Three different tests can be constructed:

Test 1 (one-sided): If

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \geq C_1(\alpha^*) \quad \text{or} \quad \inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \leq -C_1(\alpha^*),$$

then we conclude that there was a downward or upward change in parameter $\theta^{(i)}$ (respectively) along the sequence X_0, X_1, \dots, X_n .

Test 2 (two-sided): If

$$\sup_{0 \leq t \leq 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| \geq C_2(\alpha^*),$$

then we conclude that there was a change in parameter $\theta^{(i)}$ along the sequence X_0, X_1, \dots, X_n .

The third test is designed against the so-called epidemic alternative (see Csörgő & Horváth, 1997, 1.7.4) and is included for the sake of completeness, following Gombay (2008). No results will be proved under this alternative hypothesis.

Test 3: If

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) - \inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) \geq C_3(\alpha^*),$$

then we conclude that there was a temporary change in parameter $\theta^{(i)}$ along the sequence X_0, X_1, \dots, X_n .

Critical values are obtained from the limit distributions in Corollary 3.2, namely, from the identities

$$\begin{aligned} \mathbb{P} \left(\sup_{0 \leq t \leq 1} B(t) \geq x \right) &= e^{-2x^2}, \quad x \geq 0, \\ \mathbb{P} \left(\sup_{0 \leq t \leq 1} |B(t)| \geq x \right) &= 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 x^2}, \quad x \geq 0, \\ \mathbb{P} \left(\sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t) \geq x \right) &= 1 - 2 \sum_{k=1}^{\infty} (4k^2 x^2 - 1) e^{-2k^2 x^2}, \end{aligned}$$

respectively, where $(B(t))_{0 \leq t \leq 1}$ is a Brownian bridge (for the third relationship, see Kuiper, 1960).

4 The test under the alternative hypothesis

In this section we will state the two main results of this paper which extend the results of Hušková, Prášková, and Steinebach (2007) and Gombay (2008) to the INAR(p) process.

4.1 Consistency of the test

The following theorem, the analogue of Theorem 3.1 in Hušková et al. (2007), describes the behaviour of the maximum of the test process if a change occurs in the mean of the innovation. An immediate consequence of the theorem is that the maximum of the process tends to infinity stochastically as $n \rightarrow \infty$, which suffices for the weak consistency of the proposed test.

Theorem 4.1. *Suppose that H_A holds with $\tau = \max(\lfloor n\rho \rfloor, 1)$, $\rho \in (0, 1)$, and the change is only in the innovation mean, namely, the innovation mean changes from μ' to μ'' , where $\mu' > \mu'' > 0$. Suppose that $E(X_0^6) < \infty$, \dots , $E(X_{-p+1}^6) < \infty$, $E(\varepsilon_1^6) < \infty$, $E(\varepsilon_{\tau+1}^6) < \infty$, $\alpha_1 + \dots + \alpha_p < 1$ and $\mu > 0$ hold both before and after the change. Assume that $\alpha_1 + \dots + \alpha_p > 0$ or $\sigma^2 > 0$. Then for any $\gamma \in (0, \frac{1}{4})$ we have*

$$\max_{1 \leq k \leq n} \sum_{j=1}^k \widehat{M}_j^{(n)} = n\psi + O_P(n^{1-\gamma}) \quad \text{as } n \rightarrow \infty,$$

with

$$\psi := \rho(1 - \rho)(\mu' - \mu'')\mathbf{e}_{p+1}^\top \mathbf{C}' \widetilde{\mathbf{Q}}^{-1} \mathbf{C}' \mathbf{e}_{p+1} > 0,$$

where \mathbf{e}_{p+1} is the $p+1$ -st ($p+1$ -dimensional) unit vector,

$$\mathbf{C}' := E \left(\begin{bmatrix} \widetilde{\mathbf{X}}' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}}' \\ 1 \end{bmatrix}^\top \right), \quad \mathbf{C}'' := E \left(\begin{bmatrix} \widetilde{\mathbf{X}}'' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}}'' \\ 1 \end{bmatrix}^\top \right),$$

where the distributions of

$$\widetilde{\mathbf{X}}' = \begin{bmatrix} \widetilde{X}'_0 \\ \vdots \\ \widetilde{X}'_{-p+1} \end{bmatrix}, \quad \widetilde{\mathbf{X}}'' = \begin{bmatrix} \widetilde{X}''_0 \\ \vdots \\ \widetilde{X}''_{-p+1} \end{bmatrix}$$

are the unique stationary distributions of the Markov chains $(\mathbf{X}_k)_{0 \leq k \leq \lfloor n\rho \rfloor}$ and $(\mathbf{X}_k)_{\lfloor n\rho \rfloor + 1 \leq k \leq n}$, respectively, and

$$\widetilde{\mathbf{Q}} := \rho \mathbf{C}' + (1 - \rho) \mathbf{C}''.$$

Remark. Theorem 4.1 can be reformulated for a change in the other parameters as well. In other cases we have to investigate the maximum or minimum of $\sum_{j=1}^k \widehat{M}_j^{(n)} X_{j-q}$ for some $1 \leq q \leq p$. This will be explored in more detail in B.5. One warning is given here for emphasis, although it will be repeated later on, namely, the one-sided test should only be used if we are certain that only one parameter of the process has changed.

4.2 Estimation of the change point

Based on the score vector analogy described in Appendix A, the estimator of τ is

$$\hat{\tau}_n := \min \left\{ k \in \{1, \dots, n\} : \sum_{j=1}^k \widehat{M}_j^{(n)} = \max_{1 \leq m \leq n} \sum_{j=1}^m \widehat{M}_j^{(n)} \right\} \quad (4.1)$$

for the downward one-sided test,

$$\hat{\tau}_n = \min \left\{ k \in \{1, \dots, n\} : \sum_{j=1}^k \widehat{M}_j^{(n)} = \min_{1 \leq m \leq n} \sum_{j=1}^m \widehat{M}_j^{(n)} \right\} \quad (4.2)$$

for the upward one-sided test, and

$$\hat{\tau}_n = \min \left\{ k \in \{1, \dots, n\} : \left| \sum_{j=1}^k \widehat{M}_j^{(n)} \right| = \max_{1 \leq m \leq n} \left| \sum_{j=1}^m \widehat{M}_j^{(n)} \right| \right\} \quad (4.3)$$

for the two-sided test.

If there is a change in the coefficient α_i , then the estimator of τ is based on the process $\sum_{j=1}^k \widehat{M}_j^{(n)} X_{j-i}$, $k = 1, \dots, n$.

Theorem 4.2. *Under the assumptions of Theorem 4.1, we have*

$$\hat{\tau}_n - \lfloor n\rho \rfloor = O_P(1) \quad \text{as } n \rightarrow \infty,$$

where $\hat{\tau}_n$ is defined by (4.1).

Remark. We used the change-point estimator from (4.1) because in Theorem 4.1, the mean μ changes downward. If it changed upward, then we should apply the estimator from (4.2) because it corresponds to the appropriate one-sided test. The estimator from (4.3) corresponds to the two-sided test and satisfies the statement of Theorem 4.2 regardless of the direction of the change. The proof of this is analogous to the proof of Theorem 4.2 but even more technical, therefore it will be omitted.

Remark. Similarly to Theorem 4.1, Theorem 4.2 holds for a change in other parameters as well. We also note that the result is slightly stronger than the similar Proposition 3.1 in Hušková et al. (2007). Similar results are valid for change in a location parameter (see Csörgő & Horváth, 1997), and in these cases the limit distribution is nondegenerate. Therefore we can conjecture that Theorem 4.2 cannot be improved upon in terms of convergence rate.

Theorem 4.1 and Theorem 4.2 are proved in Appendix B. If we would like to prove different versions of these theorems (e.g., if we postulate a change in a parameter other than μ), we need to slightly modify the proofs in several points. Any points where these modifications are not straightforward will be indicated in B.5.

5 Illustration

Now we provide two real data examples of the use of our method. Since our model includes initial values, the series were not investigated in their full length, but the first p values were taken as the initial values X_{-p+1}, \dots, X_0 .

Our first example is the dataset of monthly polio cases in the US, as reported by the Centers for Disease Control and Prevention. It is available online at Hyndman (n.d.) and is 166 long. In Kang and Lee (2009) the authors found a significant decreasing trend in this series, while in Davis and Wu (2009) and Davis et al. (2000) the trend was found insignificant. It is widely agreed (see also Silva, 2005) that the underlying process is first-order, which is also supported by the partial autocorrelation function. Therefore we treated it as an INAR(1) process and calculated the CLS estimates given by (2.2). They were $\hat{\alpha}_1 = 0.30646$ and $\hat{\mu} = 0.94091$. The maximum of the absolute value of $\widehat{\mathcal{M}}_{166}^{(1)}$ was 1.2647 and the maximum of the absolute value of $\widehat{\mathcal{M}}_{166}^{(2)}$ was 1.1232. Applying the two-sided test simultaneously to the two parameters and requiring an overall significance level of 0.05, the critical value for each component is 1.48 (the individual significance levels are $1 - \sqrt{0.95} \cong 0.0253$), therefore, the null hypothesis is not rejected.

Our second example is a dataset of public drunkenness intakes in Minneapolis, also accessible at Hyndman (n.d.). This dataset is 139 long. After an examination of the partial autocorrelation function a seasonal INAR(12) model seems a rational choice, but with the assumption that only α_1 and α_{12} are nonzero (for another similar calculation, see the real data section in Barczy et al., 2011). The estimates are

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_{12} \\ \hat{\mu} \end{bmatrix} = \left(\sum_{k=1}^n \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix} \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix}^\top \right)^{-1} \sum_{k=1}^n X_k \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.8154 \\ 0.1419 \\ 9.6944 \end{bmatrix}.$$

The maxima of the absolute values of the respective components of $\widehat{\mathcal{M}}_n$ are 2.0333, 1.3497 and 1.5788. A comparison with the critical value of 1.545 (individual significance of approximately 0.017) results in the rejection of the null hypothesis. Based on $\left(\sum_{i=1}^k \widehat{M}_k^{(n)} X_{k-1} \right)_{k=1}^{139}$ our estimate for the change point is 41 (i.e., the 53rd entry in the original series). Repeating the procedure for the series before and after the change,

the null hypothesis is accepted for both of them. For the series after the change, the CLS estimate of α_{12} is negative but an inspection of the partial autocorrelation function reveals that this series is more appropriately modeled as an INAR(1) process, for which the parameter estimates are $\hat{\alpha}_1 = 0.8915$ and $\hat{\mu} = 24.8429$ and the null hypothesis is accepted.

A The process under the null hypothesis

In this section we will give some results for the process under the null hypothesis, including a lemma for ergodic convergence rate. The conditional least squares estimators will be calculated and we will give details on the calculation of the test process (2.7).

A.1 Regression equations

The INAR(p) process is formally analogous to the AR(p) process. To exploit this analogy we need to state several regression equations for the process. First we write the equivalent of (1.1) for the vector-valued process $(\mathbf{X}_k)_{k \in \mathbb{N}}$.

$$\mathbf{X}_k = \sum_{i=1}^p \sum_{j=1}^{X_{k-i}} \boldsymbol{\xi}_{i,k,j} + \boldsymbol{\varepsilon}_k, \quad (\text{A.1})$$

where

$$\boldsymbol{\xi}_{1,k,j} = \begin{bmatrix} \xi_{1,k,j} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{\xi}_{2,k,j} = \begin{bmatrix} \xi_{2,k,j} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \boldsymbol{\xi}_{p,k,j} = \begin{bmatrix} \xi_{p,k,j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{\varepsilon}_k = \begin{bmatrix} \varepsilon_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This form makes it even more apparent that the INAR(p) process is a special multitype branching process with immigration. According to standard literature (see, e.g., Quine, 1970), if the matrix

$$\mathbf{A} := \begin{bmatrix} \mathbb{E}(\boldsymbol{\xi}_{1,1,1}) & \dots & \mathbb{E}(\boldsymbol{\xi}_{p,1,1}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (\text{A.2})$$

is primitive (i.e., some power of it is elementwise positive), the ergodicity of the process depends only on the

spectral radius $\rho(A)$, and the process is ergodic if $\rho(A) < 1$. In Barczy et al. (2011), (2.7), it is shown that this is equivalent to the condition that $\alpha_1 + \dots + \alpha_p < 1$.

Recalling the M_k martingale differences from (2.1) we can write

$$\mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} + (\mu + M_k)\mathbf{e}_1, \quad (\text{A.3})$$

where \mathbf{e}_1 is the first unit vector. Based on (A.3) we obtain

$$\begin{aligned} \mathbf{X}_k^{\otimes 2} &= (\mathbf{A}\mathbf{X}_{k-1})^{\otimes 2} + ((\mu + M_k)\mathbf{e}_1)^{\otimes 2} + (\mathbf{A}\mathbf{X}_{k-1}) \otimes ((\mu + M_k)\mathbf{e}_1) \\ &\quad + ((\mu + M_k)\mathbf{e}_1) \otimes (\mathbf{A}\mathbf{X}_{k-1}) \\ &= \mathbf{A}^{\otimes 2} \mathbf{X}_{k-1}^{\otimes 2} + (\mu + M_k)^2 \mathbf{e}_1^{\otimes 2} + (\mu + M_k)(\mathbf{A}\mathbf{X}_{k-1}) \otimes \mathbf{e}_1 \\ &\quad + (\mu + M_k)\mathbf{e}_1 \otimes (\mathbf{A}\mathbf{X}_{k-1}), \end{aligned} \quad (\text{A.4})$$

where \otimes denotes Kronecker product of matrices.

A.2 Ergodicity

Under the null hypothesis let us denote by $\widetilde{\mathbf{X}}$ a random vector with the unique stationary distribution of $(\mathbf{X}_k)_{k \in \mathbb{N}}$. Because our process is ergodic, we can apply the ergodic theorem. In its well-known form it states that if $\mathbb{E}(|g(\widetilde{\mathbf{X}})|) < \infty$ for some function g , then

$$\frac{1}{n} \sum_{k=1}^n g(\mathbf{X}_k) \xrightarrow{\text{a.s.}} \mathbb{E}(g(\widetilde{\mathbf{X}})). \quad (\text{A.5})$$

This is, for example, Theorem 2 in I.15. in Chung (1960). However, instead of the convergence of averages, we will frequently require the convergence of expectations, i.e.,

$$\mathbb{E}(\mathbf{X}_k^{\otimes \beta}) \rightarrow \mathbb{E}(\widetilde{\mathbf{X}}^{\otimes \beta}), \quad \beta \in \mathbb{N}, \quad (\text{A.6})$$

whenever the right hand side is finite. This is Theorem 14.0.1 in Meyn and Tweedie (2009). The result in (A.6) also implies convergence of any component of the matrices. Under the alternative hypothesis we will additionally apply

$$\sum_{x \in \mathbb{N}_0^p} |\mathbb{P}(\mathbf{X}_n = x) - \mathbb{P}(\widetilde{\mathbf{X}} = x)| \rightarrow 0. \quad (\text{A.7})$$

This result can be found in Orey (1971, Theorem 4.3) or in Meyn and Tweedie (2009, Theorem 13.1.2).

The convergence rate in (A.6) can be estimated by the following lemma.

Lemma A.1. *There is a constant $\pi \in (0, 1)$ such that*

$$\|E(\mathbf{X}_k) - E(\widetilde{\mathbf{X}})\| = O(\pi^k), \quad \|E(\mathbf{X}_k^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2})\| = O(\pi^k).$$

Proof. We use (A.3) to conclude that

$$E(\mathbf{X}_k) = \mathbf{A} E(\mathbf{X}_{k-1}) + \mu \mathbf{e}_1.$$

Taking the limits as $k \rightarrow \infty$ we have

$$E(\widetilde{\mathbf{X}}) = \mathbf{A} E(\widetilde{\mathbf{X}}) + \mu \mathbf{e}_1,$$

hence

$$E(\mathbf{X}_k) - E(\widetilde{\mathbf{X}}) = \mathbf{A}(E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}})). \quad (\text{A.8})$$

Similarly, from (A.4) and $E(M_k^2 | \mathcal{F}_{k-1}) = \boldsymbol{\alpha}_2^\top \mathbf{X}_{k-1} + \sigma^2$ we have

$$\begin{aligned} E(\mathbf{X}_k^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2}) &= \mathbf{A}^{\otimes 2}(E(\mathbf{X}_{k-1}^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2})) + (\boldsymbol{\alpha}_2^\top (E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}}))) \mathbf{e}_1^{\otimes 2} \\ &\quad + \mu(\mathbf{A}(E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}}))) \otimes \mathbf{e}_1 + \mu \mathbf{e}_1 \otimes (\mathbf{A}(E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}}))) \\ &= \mathbf{A}^{\otimes 2}(E(\mathbf{X}_{k-1}^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2})) + \mathbf{e}_1^{\otimes 2} \boldsymbol{\alpha}_2^\top (E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}})) \\ &\quad + \mu(\mathbf{A} \otimes \mathbf{e}_1)(E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}})) + \mu(\mathbf{e}_1 \otimes \mathbf{A})(E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}})). \end{aligned} \quad (\text{A.9})$$

Here we used the fact that for any $c \in \mathbb{R}$ and real vector \mathbf{v} we have $c\mathbf{v} = \mathbf{v}c$, where the second multiplication is a proper matrix product. Furthermore, we used the following property of the Kronecker product: for any matrices A, B, C, D we have $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$, specifically, if C is a column vector, $(AB) \otimes C = (AB) \otimes (C \cdot 1) = (A \otimes C)(B \otimes 1) = (A \otimes C)B$ (this identity can also be used when the first factor consists of a single factor instead of the second). Hence,

$$\begin{bmatrix} E(\mathbf{X}_k) - E(\widetilde{\mathbf{X}}) \\ E(\mathbf{X}_k^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2}) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{e}_1^{\otimes 2} \boldsymbol{\alpha}_2^\top + \mu(\mathbf{A} \otimes \mathbf{e}_1) + \mu(\mathbf{e}_1 \otimes \mathbf{A}) & \mathbf{A}^{\otimes 2} \end{bmatrix} \begin{bmatrix} E(\mathbf{X}_{k-1}) - E(\widetilde{\mathbf{X}}) \\ E(\mathbf{X}_{k-1}^{\otimes 2}) - E(\widetilde{\mathbf{X}}^{\otimes 2}) \end{bmatrix}.$$

Let us denote the multiplying matrix on the right hand side by \mathbf{D} . We note that \mathbf{D} is block lower triangular and that due to the properties of the Kronecker product, $\rho(\mathbf{A}^{\otimes 2}) = (\rho(\mathbf{A}))^2 < \rho(\mathbf{A})$ (here and throughout the paper, ρ denotes the spectral radius of a matrix). From these it is clear that $\rho(\mathbf{D}) = \rho(\mathbf{A}) < 1$. It is well-known that then there exists an induced matrix norm $\|\cdot\|_*$ for which $\rho(\mathbf{A}) < \|\mathbf{A}\|_* < 1$. This,

and the equivalence of vector norms suffice for the proof. \square

Remark. The finiteness of the respective moments of the stationary distribution can be derived using the same approach as in the proof of formulae (2.2.3), (2.2.4) and (2.2.10) in Barczy et al. (2011). We note that the stationary distribution has exactly as many finite moments as the innovation distribution and the initial distributions have in common, because the Bernoulli distribution is bounded and therefore all of its moments are finite.

A.3 The variance of partial sums of the process

The following lemma will be used for estimations in the next section and shows that while the values of the process are not independent, their dependence is very weak in the sense that the variance of the partial sums up to time n is only linearly increasing with n .

Lemma A.2. *In a stable INAR(p) model under H_0*

- (i) $\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = O(n),$
- (ii) $\text{Var}(X_1 X_{1-q} + X_2 X_{2-q} + \dots + X_n X_{n-q}) = \sum_{i,j=1}^n \text{Cov}(X_i X_{i-q}, X_j X_{j-q}) = O(n)$ for all $0 \leq q \leq p-1.$

Proof. Although the lemma is stated for the process $(X_n)_{n \in \mathbb{N}}$, calculations will require that we investigate the process $(\mathbf{X}_n)_{n \in \mathbb{N}}$. Therefore, we will prove the following statements:

$$\|\text{Var}(\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n)\| = O(n) \quad (\text{A.10})$$

in the place of (i) and

$$\|\text{Var}(\mathbf{X}_1^{\otimes 2} + \mathbf{X}_2^{\otimes 2} + \dots + \mathbf{X}_n^{\otimes 2})\| = O(n) \quad (\text{A.11})$$

in the place of (ii).

First we will prove (A.10). (A.6) implies that $(\|\text{Var}(\mathbf{X}_i)\|)_{i \in \mathbb{N}}$ is a (convergent and hence) bounded series. Let us denote its upper bound by U_1 . This observation means that we will only need to investigate the sum $\sum_{i,j=1}^n \text{Cov}(\mathbf{X}_i, \mathbf{X}_j) - \sum_{i=1}^n \text{Cov}(\mathbf{X}_i, \mathbf{X}_i)$, since the latter sum is clearly $O(n)$.

First we will use (A.3). It is immediate from (A.3) that

$$\mathbf{X}_k - \mathbb{E}(\mathbf{X}_k) = \mathbf{A}(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})) + M_k \mathbf{e}_1.$$

Let us now fix $1 \leq i < j$ and write

$$\begin{aligned}
\text{Cov}(\mathbf{X}_i, \mathbf{X}_j) &= \mathbb{E}[(\mathbf{X}_i - \mathbb{E}(\mathbf{X}_i))(\mathbf{X}_j - \mathbb{E}(\mathbf{X}_j))^\top] \\
&= \mathbb{E}[(\mathbf{X}_i - \mathbb{E}(\mathbf{X}_i))\mathbb{E}(\mathbf{X}_j - \mathbb{E}(\mathbf{X}_j)|\mathcal{F}_{j-1})^\top] \\
&= \mathbb{E}[(\mathbf{X}_i - \mathbb{E}(\mathbf{X}_i))(\mathbf{A}(\mathbf{X}_{j-1} - \mathbb{E}(\mathbf{X}_{j-1})))^\top] = \text{Cov}(\mathbf{X}_i, \mathbf{X}_{j-1})\mathbf{A}^\top.
\end{aligned} \tag{A.12}$$

If we perform the calculations for the case $1 \leq j < i$ as well, we can see that, after $|j - i|$ iterations,

$$\text{Cov}(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{A}^{(i-j)_+} \text{Var}(\mathbf{X}_{\min(i,j)})(\mathbf{A}^\top)^{(j-i)_+}. \tag{A.13}$$

Because $\rho(\mathbf{A}) < 1$, there exists a matrix norm $\|\cdot\|_*$ for which $\rho(\mathbf{A}) < \|\mathbf{A}\|_* =: \pi < 1$. With this norm, (A.13), and the boundedness of $(\text{Var}(\mathbf{X}_i))_{i \in \mathbb{N}}$ we can establish

$$\|\text{Cov}(\mathbf{X}_i, \mathbf{X}_j)\|_* = O(\pi^{|i-j|}),$$

which yields (A.10) immediately.

For (A.11) our reasoning will be very similar, although with more tedious calculations. First we note that (A.6) implies boundedness for $\|\text{Var}(\mathbf{X}_i^{\otimes 2})\|$ also. From (A.4) we have

$$\begin{aligned}
\mathbb{E}(\mathbf{X}_k^{\otimes 2} - \mathbb{E}(\mathbf{X}_k^{\otimes 2})|\mathcal{F}_{k-1}) &= \mathbf{A}^{\otimes 2}(\mathbf{X}_{k-1}^{\otimes 2} - \mathbb{E}(\mathbf{X}_{k-1}^{\otimes 2})) + \boldsymbol{\alpha}_2^\top(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1}))\mathbf{e}_1^{\otimes 2} \\
&\quad + \mu[\mathbf{A}(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1}))] \otimes \mathbf{e}_1 + \mu\mathbf{e}_1 \otimes [\mathbf{A}(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1}))] \\
&= \mathbf{A}^{\otimes 2}(\mathbf{X}_{k-1}^{\otimes 2} - \mathbb{E}(\mathbf{X}_{k-1}^{\otimes 2})) + \mathbf{e}_1^{\otimes 2}\boldsymbol{\alpha}_2^\top(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})) \\
&\quad + \mu[\mathbf{A} \otimes \mathbf{e}_1](\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})) \otimes + \mu(\mathbf{e}_1 \otimes \mathbf{A})(\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})),
\end{aligned}$$

analogously to (A.9). Now, similarly to (A.12) we get, for $1 \leq i < j$,

$$\begin{aligned}
\text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j^{\otimes 2}) &= \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1}^{\otimes 2})(\mathbf{A}^{\otimes 2})^\top + \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1})\boldsymbol{\alpha}_2(\mathbf{e}_1^{\otimes 2})^\top \\
&\quad + \mu \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1})(\mathbf{A} \otimes \mathbf{e}_1)^\top + \mu \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1})(\mathbf{e}_1 \otimes \mathbf{A})^\top.
\end{aligned}$$

Here

$$\text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1}) := \mathbb{E}[(\mathbf{X}_i^{\otimes 2} - \mathbb{E}(\mathbf{X}_i^{\otimes 2}))(\mathbf{X}_{j-1} - \mathbb{E}(\mathbf{X}_{j-1}))^\top],$$

a $p^2 \times p$ matrix. Also similarly to (A.12) we have

$$\text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j) = \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1})\mathbf{A}^\top.$$

Summarizing, we get the following regression:

$$\begin{bmatrix} \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j)^\top \\ \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j^{\otimes 2})^\top \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\otimes 2} & \mathbf{0} \\ \mathbf{e}_1^{\otimes 2} \boldsymbol{\alpha}_2^\top + \mu(\mathbf{A} \otimes \mathbf{e}_1) + \mu(\mathbf{e}_1 \otimes \mathbf{A}) & \mathbf{A} \end{bmatrix} \begin{bmatrix} \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1})^\top \\ \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_{j-1}^{\otimes 2})^\top \end{bmatrix}. \quad (\text{A.14})$$

Note that the multiplying matrix on the right hand side is just \mathbf{D} from the proof of Lemma A.1. Now, similarly to (A.13), we have

$$\begin{bmatrix} \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j)^\top \\ \text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_j^{\otimes 2})^\top \end{bmatrix} = \mathbf{D}^{(j-i)_+} \begin{bmatrix} \text{Cov}(\mathbf{X}_{\min(i,j)}^{\otimes 2}, \mathbf{X}_{\min(i,j)})^\top \\ \text{Cov}(\mathbf{X}_{\min(i,j)}^{\otimes 2}, \mathbf{X}_{\min(i,j)}^{\otimes 2})^\top \end{bmatrix} (\mathbf{D}^\top)^{(i-j)_+}. \quad (\text{A.15})$$

Now we only need to note that $(\text{Cov}(\mathbf{X}_i^{\otimes 2}, \mathbf{X}_i))_{i \in \mathbb{N}}$ is a bounded sequence (due to (A.6)), and we can finish the proof of (A.11) in the same way as (A.10). \square

A.4 Strong consistency of the estimates

In this subsection we want to show with our notations the fact, proven already by Du and Li (1991) for $\boldsymbol{\theta}$, that the estimates for the coefficients and the mean and variance of the innovation given in Section 2 are strongly consistent, i.e.,

$$\widehat{\boldsymbol{\theta}}^{(n)} \xrightarrow{\text{a.s.}} \boldsymbol{\theta} \quad \text{and} \quad \widehat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \sigma^2. \quad (\text{A.16})$$

First of all, it is a matter of simple calculations (see C.2) and a straightforward application of (A.5) that, provided that the second moment of ε_1 exists,

$$\frac{1}{n} \sum_{k=1}^n M_k^2 \rightarrow \boldsymbol{\alpha}_2 \mathbb{E}(\widetilde{\mathbf{X}}) + \sigma^2, \quad (\text{A.17})$$

with $\boldsymbol{\alpha}_2$ given in (2.6). Hence, taking the limits of the expectations in (A.3) and (A.4) we have

$$\mathbb{E}(\widetilde{\mathbf{X}}) = \mathbf{A} \mathbb{E}(\widetilde{\mathbf{X}}) + \mu \quad \text{and} \quad \mathbb{E}(\widetilde{\mathbf{X}}^{\otimes 2}) = \mathbf{A}^{\otimes 2} \mathbb{E}(\widetilde{\mathbf{X}}^{\otimes 2}) + (\mu^2 + \boldsymbol{\alpha}_2^\top \mathbb{E}(\widetilde{\mathbf{X}}) + \sigma^2) + \mu(\mathbf{e}_1 \otimes \mathbf{A} \mathbb{E}(\widetilde{\mathbf{X}}) + \mathbf{A} \mathbb{E}(\widetilde{\mathbf{X}}) \otimes \mathbf{e}_1).$$

From this we can deduce, using (A.6) and an argument from the proof of Lemma B.1,

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{(n)} &= \left(\frac{\mathbf{Q}_n}{n}\right)^{-1} \frac{1}{n} \sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \\ &\rightarrow \mathbb{E} \left(\begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right)^{-1} \mathbb{E} \left((\boldsymbol{\alpha}^\top \widetilde{\mathbf{X}} + \mu) \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \right) = \mathbb{E} \left(\begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right)^{-1} \mathbb{E} \left(\begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right) \boldsymbol{\theta} = \boldsymbol{\theta}.\end{aligned}$$

A similar result can be derived for the estimate of σ^2 . By recalling (A.17) and computing the strong limit of the other summands in (2.4), we obtain the strong consistency of $\overline{\sigma}_n^2$ immediately. The same reasoning shows that if the second moment of the stationary distribution is finite (in this case we already know that $\widehat{\boldsymbol{\theta}}^{(n)}$ is a consistent estimator), then the limits of the estimators $\overline{\sigma}_n^2$ and $\widehat{\sigma}_n^2$ are the same almost surely; hence, the strong consistency of $\widehat{\sigma}_n^2$ is established.

A.5 Proof of Theorem 3.1

We recall that the theorem was stated under the null hypothesis, so we will assume it to hold. Let

$$\mathcal{Z}_n(t) := \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \mathbf{Z}_k, \quad t \in [0, 1], \quad \mathbf{Z}_k := M_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}, \quad k = 1, 2, \dots$$

The proof will be based on the following theorem.

Theorem A.3. *Under the assumptions of Theorem 3.1*

$$\mathcal{Z}_n \xrightarrow{\mathcal{D}} \mathbf{I}^{1/2} \mathcal{W}, \quad n \rightarrow \infty,$$

where $(\mathcal{W}(t))_{0 \leq t \leq 1}$ is a $(p+1)$ -dimensional standard Wiener process and

$$\mathbf{I} := \mathbb{E} \left((\boldsymbol{\alpha}_2^\top \widetilde{\mathbf{X}} + \sigma^2) \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right). \quad (\text{A.18})$$

Proof. By (A.5) we have $n^{-1} \mathbf{I}_n \xrightarrow{\text{a.s.}} \mathbf{I}$, and since $\widehat{\sigma}_n^2$ and $\widehat{\boldsymbol{\theta}}_n$ are strongly consistent estimators, therefore we have $n^{-1} \widehat{\mathbf{I}}_n \xrightarrow{\text{a.s.}} \mathbf{I}$ as well. We will use the martingale central limit theorem for the martingale differences

$\frac{1}{\sqrt{n}}\mathbf{Z}_k$, $n \in \mathbb{N}$, $k = 1, 2, \dots, n$. To compute the variance function, we write

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}(\mathbf{Z}_k \mathbf{Z}_k^\top | \mathcal{F}_{k-1}) &= \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top \\ &= \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} (\boldsymbol{\alpha}_2^\top \mathbf{X}_{k-1} + \sigma^2) \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top \\ &\xrightarrow{\text{a.s.}} t \mathbb{E} \left((\boldsymbol{\alpha}_2^\top \widetilde{\mathbf{X}} + \sigma^2) \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right) = t\mathbf{I}. \end{aligned}$$

It remains to check the so-called conditional Lindeberg condition:

$$\begin{aligned} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E} \left(\left\| \frac{1}{\sqrt{n}} \mathbf{Z}_k \right\|^2 \chi_{\{\|n^{-1/2} \mathbf{Z}_k\| > \delta\}} \middle| \mathcal{F}_{k-1} \right) &\leq \frac{1}{\delta^2} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E} \left(\left\| \frac{1}{\sqrt{n}} \mathbf{Z}_k \right\|^4 \middle| \mathcal{F}_{k-1} \right) \\ &= \frac{1}{\delta^2 n^2} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}(M_k^4 | \mathcal{F}_{k-1}) (X_{k-1}^2 + \dots + X_{k-p}^2 + 1)^2 \\ &= \frac{1}{\delta^2 n^2} \sum_{k=1}^{\lfloor nt \rfloor} P(\mathbf{X}_{k-1}), \end{aligned}$$

where P is a polynomial of degree six, because $\mathbb{E}(M_k^4 | \mathcal{F}_{k-1})$ is a second-degree polynomial of \mathbf{X}_{k-1} (this is detailed in C.2). The sixth moment of the stationary distribution is finite due to the assumptions, hence (A.5) implies

$$\frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} P(\mathbf{X}_{k-1}) \xrightarrow{\text{a.s.}} \mathbb{E}(P(\widetilde{\mathbf{X}})) < \infty.$$

This means

$$\frac{1}{\delta^2 n^2} \sum_{k=1}^{\lfloor nt \rfloor} P(\mathbf{X}_{k-1}) \xrightarrow{\text{a.s.}} 0,$$

implying Lindeberg's condition. All the conditions of the martingale central limit theorem have been checked; the proof is therefore complete. \square

Having proved Theorem A.3, we now turn to the proof of Theorem 3.1. For this, let us introduce the notation

$$\widehat{\mathbf{Z}}_k^{(n)} := \widehat{M}_k^{(n)} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}, \quad k = 1, 2, \dots$$

First we note that

$$\sum_{k=1}^{\lfloor nt \rfloor} \widehat{\mathbf{Z}}_k^{(n)} = \sum_{k=1}^{\lfloor nt \rfloor} \mathbf{Z}_k + \sum_{k=1}^{\lfloor nt \rfloor} (\widehat{\mathbf{Z}}_k^{(n)} - \mathbf{Z}_k) = \sum_{k=1}^{\lfloor nt \rfloor} \mathbf{Z}_k + \sum_{k=1}^{\lfloor nt \rfloor} (\widehat{M}_k^{(n)} - M_k) \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}.$$

Recalling the definitions of M_k and $\widehat{M}_k^{(n)}$,

$$\widehat{M}_k^{(n)} - M_k = \left(X_k - \widehat{\boldsymbol{\theta}}_n^\top \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right) - \left(X_k - \boldsymbol{\theta}^\top \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right) = (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}.$$

We have

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} &= \mathbf{Q}_n^{-1} \left(\sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right) - \boldsymbol{\theta} \\ &= \mathbf{Q}_n^{-1} \left(\sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} - \sum_{k=1}^n \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top \boldsymbol{\theta}^\top \right) \\ &= \mathbf{Q}_n^{-1} \left(\sum_{k=1}^n M_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right), \end{aligned} \tag{A.19}$$

hence

$$\widehat{\mathbf{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{\mathbf{Z}}_k^{(n)} = \sqrt{n} \widehat{\mathbf{I}}_n^{-1/2} \left(\sum_{k=1}^{\lfloor nt \rfloor} \frac{1}{\sqrt{n}} \mathbf{Z}_k - \mathbf{Q}_{\lfloor nt \rfloor} \mathbf{Q}_n^{-1} \sum_{k=1}^n \frac{1}{\sqrt{n}} \mathbf{Z}_k \right). \tag{A.20}$$

In the next step we notice that according to (A.5) we have

$$\mathbf{Q}_{\lfloor nt \rfloor} \mathbf{Q}_n^{-1} = \frac{\lfloor nt \rfloor}{n} \left(\frac{1}{\lfloor nt \rfloor} \mathbf{Q}_{\lfloor nt \rfloor} \right) \left(\frac{1}{n} \mathbf{Q}_n \right)^{-1} \xrightarrow{\text{a.s.}} t \widetilde{\mathbf{Q}}_{\text{H}_0} \widetilde{\mathbf{Q}}_{\text{H}_0}^{-1} = t \mathbf{E}_{p+1} \quad \forall t \in [0, 1],$$

where \mathbf{E}_{p+1} is the $p+1$ -dimensional identity matrix and

$$\widetilde{\mathbf{Q}}_{\text{H}_0} := \mathbb{E} \left(\begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{X}} \\ 1 \end{bmatrix}^\top \right).$$

Now we apply (A.20), Theorem A.3, and

$$\sqrt{n} \widehat{\mathbf{I}}_n^{-1/2} \xrightarrow{\text{a.s.}} \mathbf{I}^{-1/2},$$

to conclude that

$$\left(\widehat{\mathbf{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{\mathbf{Z}}_k^{(n)} \right)_{t \in [0,1]} \xrightarrow{\mathcal{D}} (\mathbf{W}(t) - t\mathbf{W}(1))_{t \in [0,1]}. \quad \square$$

B The process under the alternative hypothesis

While in Theorem 3.1 we were able to consider longer and longer samples taken from the same process, this approach has to be modified for the alternative hypothesis. More precisely, we have to consider a series of time-inhomogeneous $\text{INAR}(p)$ processes, where the n -th one has a point of change at $\lfloor n\rho \rfloor$ (we will suppress this in the notation for simplicity). Now, the parts of these processes before the change (i.e., $(X_i)_{i=1}^{\lfloor n\rho \rfloor}$) can be handled as a sample taken from an infinite $\text{INAR}(p)$ process (at least in distribution), but this is not true for the second part (i.e., $(X_i)_{i \geq \lfloor n\rho \rfloor + 1}$), because the initial distribution of this process depends on n .

Therefore, for a rigorous analysis we need to refine the results of A.2.

B.1 Ergodicity under the alternative hypothesis

We have

$$\frac{1}{n - \lfloor n\rho \rfloor} \sum_{k=\lfloor n\rho \rfloor + 1}^n g(\mathbf{X}_k) \xrightarrow{\text{P}} \text{E}(g(\widetilde{\mathbf{X}}'')), \quad (\text{B.1})$$

where $g : \mathbb{N}_0^p \rightarrow \mathbb{R}$ with $\text{E}(|g(\widetilde{\mathbf{X}})|) < \infty$. Indeed, for an arbitrary $\varepsilon > 0$

$$\begin{aligned} & \text{P} \left(\left| \frac{1}{n - \lfloor n\rho \rfloor} \sum_{k=\lfloor n\rho \rfloor + 1}^n g(\mathbf{X}_k) - \text{E}(g(\widetilde{\mathbf{X}}'')) \right| > \varepsilon \right) \\ &= \sum_{x \in \mathbb{N}_0^p} \text{P} \left(\left| \frac{1}{n - \lfloor n\rho \rfloor} \sum_{k=\lfloor n\rho \rfloor + 1}^n g(\mathbf{X}_k) - \text{E}(g(\widetilde{\mathbf{X}}'')) \right| > \varepsilon \mid \mathbf{X}_{\lfloor n\rho \rfloor} = x \right) \text{P}(\mathbf{X}_{\lfloor n\rho \rfloor} = x), \end{aligned}$$

and

$$\text{P} \left(\left| \frac{1}{n - \lfloor n\rho \rfloor} \sum_{k=\lfloor n\rho \rfloor + 1}^n g(\mathbf{X}_k) - \text{E}(g(\widetilde{\mathbf{X}}'')) \right| > \varepsilon \mid \mathbf{X}_{\lfloor n\rho \rfloor} = x \right) \rightarrow 0$$

by the ergodic theorem for each $x \in \mathbb{N}_0^p$, additionally

$$\text{P}(\mathbf{X}_{\lfloor n\rho \rfloor} = x) \leq |\text{P}(\mathbf{X}_{\lfloor n\rho \rfloor} = x) - \text{P}(\widetilde{\mathbf{X}} = x)| + \text{P}(\widetilde{\mathbf{X}} = x),$$

and one can use (A.7). We will also apply that for all $\varepsilon > 0$ there exists ν such that

$$\| \text{E}(\mathbf{X}_{\lfloor n\rho \rfloor + k}) - \text{E}(\widetilde{\mathbf{X}}'') \| < \varepsilon \quad \text{for all } n \geq \nu \text{ and all } k \geq \nu. \quad (\text{B.2})$$

First observe that there exists $\pi'' \in (0, 1)$ such that $\|E(\mathbf{X}_{\lfloor n\rho \rfloor + k}) - E(\widetilde{\mathbf{X}}'')\| \leq (\pi'')^k \|E(\mathbf{X}_{\lfloor n\rho \rfloor}) - E(\widetilde{\mathbf{X}}'')\|$ for all $k \in \mathbb{N}$, see (A.8). Next, for all $\eta > 0$, choose ν_1 and ν_2 such that $(\pi'')^k < \eta$ for all $k \geq \nu_1$ and $\|E(\mathbf{X}_{\lfloor n\rho \rfloor}) - E(\widetilde{\mathbf{X}}')\| < \eta$ for all $n \geq \nu_2$. Hence

$$\|E(\mathbf{X}_{\lfloor n\rho \rfloor + k}) - E(\widetilde{\mathbf{X}}'')\| \leq \eta(\|E(\mathbf{X}_{\lfloor n\rho \rfloor}) - E(\widetilde{\mathbf{X}}')\| + \|E(\widetilde{\mathbf{X}}') - E(\widetilde{\mathbf{X}}'')\|) \leq \eta^2 + \eta\|E(\widetilde{\mathbf{X}}') - E(\widetilde{\mathbf{X}}'')\|.$$

B.2 Behaviour of the estimates under the alternative hypothesis

In this subsection we will investigate the weak limit of the CLS estimates under the conditions of Theorem 4.1. We will also show how the quantity ψ arises naturally if we would like to know how much the misestimation of the parameters influences our test process.

Lemma B.1. *Under the assumptions of Theorem 4.1 we have*

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \widetilde{\boldsymbol{\theta}} := \begin{bmatrix} \widetilde{\boldsymbol{\alpha}} \\ \widetilde{\mu} \end{bmatrix} := \widetilde{\mathbf{Q}}^{-1} \left(\rho \mathbf{C}' \begin{bmatrix} \boldsymbol{\alpha} \\ \mu' \end{bmatrix} + (1 - \rho) \mathbf{C}'' \begin{bmatrix} \boldsymbol{\alpha} \\ \mu'' \end{bmatrix} \right).$$

Proof. By (A.5) and (B.1) we obtain

$$\frac{1}{n} \mathbf{Q}_n = \frac{1}{n} \mathbf{Q}_{\lfloor n\rho \rfloor} + \frac{1}{n} \sum_{k=\lfloor n\rho \rfloor + 1}^n \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top \xrightarrow{P} \rho \mathbf{C}' + (1 - \rho) \mathbf{C}'' = \widetilde{\mathbf{Q}},$$

as $n \rightarrow \infty$. Moreover, we can notice that for a homogeneous model the process

$$\mathbf{U}_k := \begin{bmatrix} X_k \\ \mathbf{X}_{k-1} \end{bmatrix}, \quad k \in \mathbb{N}$$

satisfies a similar recursion to (A.1). The equivalent of the matrix \mathbf{A} can then be shown to have a spectral radius smaller than 1, and by the same reasoning as in A.1 we can conclude that $(\mathbf{U}_k)_{k \in \mathbb{N}}$ is ergodic, and we can apply (A.5). Moreover, it is clear that if $\widetilde{\mathbf{U}}$ denotes a vector with the unique stationary distribution of $(\mathbf{U}_k)_{k \in \mathbb{N}}$ then

$$\begin{bmatrix} \widetilde{U}^{(2)} \\ \widetilde{U}^{(3)} \\ \vdots \\ \widetilde{U}^{(p+1)} \end{bmatrix} \stackrel{\mathcal{D}}{=} \widetilde{\mathbf{X}}$$

and for the components of $\tilde{\mathbf{U}}$ we also have

$$\tilde{U}^{(1)} \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\tilde{U}^{(2)}} \xi_{1,j} + \cdots + \sum_{j=1}^{\tilde{U}^{(p+1)}} \xi_{p,j} + \varepsilon,$$

where $\xi_{i,j} \stackrel{\mathcal{D}}{=} \xi_{i,1,1}, i = 1, \dots, p, j \in \mathbb{N}$ and $\varepsilon \stackrel{\mathcal{D}}{=} \varepsilon_1$ such that all these variables are totally independent and also independent of $(\tilde{U}^{(2)}, \dots, \tilde{U}^{(p+1)})^\top$. Hence (A.5) implies.

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{\lfloor n\rho \rfloor} X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} &\xrightarrow{\text{P}} \rho \mathbb{E} \left(\left(\sum_{j=1}^{\tilde{X}'_0} \xi_{1,j} + \cdots + \sum_{j=1}^{\tilde{X}'_{-p+1}} \xi_{p,j} + \varepsilon' \right) \begin{bmatrix} \tilde{\mathbf{X}}' \\ 1 \end{bmatrix} \right) \\ &= \rho \mathbb{E} \left(\left(\alpha_1 \tilde{X}'_0 + \cdots + \alpha_p \tilde{X}'_{-p+1} + \mu' \right) \begin{bmatrix} \tilde{\mathbf{X}}' \\ 1 \end{bmatrix} \right) = \rho \mathbf{C}' \begin{bmatrix} \boldsymbol{\alpha} \\ \mu' \end{bmatrix} \end{aligned}$$

as $n \rightarrow \infty$ (here $\varepsilon' \stackrel{\mathcal{D}}{=} \varepsilon_1$). In a similar way, using (B.1)

$$\frac{1}{n} \sum_{k=\lfloor n\rho \rfloor + 1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \xrightarrow{\text{P}} (1 - \rho) \mathbf{C}'' \begin{bmatrix} \boldsymbol{\alpha} \\ \mu'' \end{bmatrix}. \quad \square$$

For the rate of convergence we have the following result:

Lemma B.2. *Under the conditions of Theorem 4.1 we have*

$$\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}} = \text{O}_{\text{P}}(n^{-1/2}).$$

Proof. The difference can be decomposed in the following way:

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}) &= (n^{-1} \mathbf{Q}_n)^{-1} n^{-1/2} \left[\sum_{k=1}^{\lfloor n\rho \rfloor} X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} - \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \left(\rho \mathbf{C}' \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix} \right) \right. \\ &\quad \left. + \sum_{k=\lfloor n\rho \rfloor + 1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} - \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \left((1 - \rho) \mathbf{C}'' \begin{bmatrix} \boldsymbol{\alpha}'' \\ \mu'' \end{bmatrix} \right) \right]. \end{aligned} \quad (\text{B.3})$$

The first factor converges to $\tilde{\mathbf{Q}}^{-1}$ stochastically, and will therefore be omitted from further calculations. The second factor has been split in two and only the first part will be analyzed in detail. The analysis of

the second part is completely analogous. We split the first part in the second factor in the following way:

$$\begin{aligned} & n^{-1/2} \left[\sum_{k=1}^{\lfloor n\rho \rfloor} X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} - \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \left(\rho \mathbf{C}' \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix} \right) \right] \\ &= n^{-1/2} \left(\sum_{k=1}^{\lfloor n\rho \rfloor} M'_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \right) + n^{-1/2} \left(\mathbf{Q}_{\lfloor n\rho \rfloor} - \rho \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \right) \begin{bmatrix} \boldsymbol{\alpha} \\ \mu' \end{bmatrix}. \end{aligned}$$

The first term is

$$n^{-1/2} \left(\sum_{k=1}^{\lfloor n\rho \rfloor} \mathbf{Z}'_k \right),$$

which is asymptotically normal, and therefore $\text{Op}(1)$ according to Lemma A.3 (the same reasoning applies after the change, since Lindeberg's theorem is valid for triangular arrays as well). We will decompose $n^{-1/2} \left(\mathbf{Q}_{\lfloor n\rho \rfloor} - \rho \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \right)$ in the following way:

$$\begin{aligned} n^{-1/2} \left(\mathbf{Q}_{\lfloor n\rho \rfloor} - \rho \mathbf{Q}_n \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \right) &= n^{-1/2} \left(\mathbf{Q}_{\lfloor n\rho \rfloor} - \mathbb{E}(\mathbf{Q}_{\lfloor n\rho \rfloor}) \right) + n^{-1/2} \left[\mathbb{E}(\mathbf{Q}_{\lfloor n\rho \rfloor}) - \lfloor n\rho \rfloor \mathbf{C}' \right] \\ &\quad - n^{-1/2} \{ \rho [\mathbf{Q}_n - \mathbb{E}(\mathbf{Q}_n)] \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \} - n^{-1/2} \{ \rho [\mathbb{E}(\mathbf{Q}_n) - n \tilde{\mathbf{Q}}] \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \} \\ &\quad - n^{-1/2} \{ \rho n \mathbf{C}' - \lfloor n\rho \rfloor \mathbf{C}' \}. \end{aligned}$$

The last term is deterministic and $o(1)$. We know from (A.11) that the variances of the first and third terms are bounded. Denoting the common upper bound by K we have, from Markov's inequality, for all n ,

$$\mathbb{P} \left(n^{-1} \left\| \mathbf{Q}_{\lfloor n\rho \rfloor} - \mathbb{E}(\mathbf{Q}_{\lfloor n\rho \rfloor}) \right\|^2 > a \right) < \frac{K}{a} \rightarrow 0 \text{ as } a \rightarrow \infty,$$

and similarly for the third term. Consequently, the first and third terms are $\text{Op}(1)$. Recalling Lemma A.1 we have

$$\begin{aligned} \left\| \mathbb{E}(\mathbf{Q}_{\lfloor n\rho \rfloor}) - \lfloor n\rho \rfloor \mathbf{C}' \right\| &= \left\| \sum_{k=1}^{\lfloor n\rho \rfloor} \left(\mathbb{E} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top - \mathbf{C}' \right) \right\| \\ &\leq \sum_{k=1}^{\lfloor n\rho \rfloor} \left\| \mathbb{E} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix}^\top - \mathbf{C}' \right\| \leq \sum_{k=1}^{\lfloor n\rho \rfloor} \pi^k = O(1), \end{aligned}$$

because the matrices within the sum consist entirely of the entries of $\mathbf{X}_k^{\otimes 2} - \widetilde{\mathbf{X}}^{\otimes 2}$. A similar calculation is valid for the fourth term. This implies the boundedness of the second and fourth terms. \square

B.3 Proof of Theorem 4.1

Now we turn to the proof of our main results: Theorem 4.1 and Theorem 4.2. We will use the following notations:

$$M'_k := X_k - \boldsymbol{\alpha}^\top \mathbf{X}_{k-1} - \mu', \quad \mathbf{Z}'_k := M'_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix},$$

and similarly for M''_k and \mathbf{Z}''_k . The proof will be given for the process before $\lfloor n\rho \rfloor$ in detail. The analysis of the process after $\lfloor n\rho \rfloor$ can be handled analogously. In the proof we will rely repeatedly on ideas from Hušková et al. (2007).

The first step is the following decomposition for $k < \lfloor n\rho \rfloor$:

$$\begin{aligned} \widehat{M}_k^{(n)} &= M'_k + [(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \mathbb{E}(\mathbf{X}_{k-1}) + (\mu' - \widetilde{\mu})] \\ &\quad + (\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}_n)^\top (\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})) \\ &\quad + [(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}_n)^\top \mathbb{E}(\mathbf{X}_{k-1}) + (\widetilde{\mu} - \widehat{\mu}_n)]. \end{aligned} \quad (\text{B.4})$$

For $k \geq \lfloor n\rho \rfloor$, the quantities M'_k and μ' have to be replaced with M''_k and μ'' , respectively. Based on (B.4) we have:

$$\begin{aligned} \left| \max_{1 \leq k \leq n} \left(\sum_{i=1}^k \widehat{M}_i^{(n)} - n\psi \right) \right| &\leq \max_{1 \leq k \leq n} \left| \sum_{i=1}^{\lfloor n\rho \rfloor \wedge k} M'_i + \sum_{i=\lfloor n\rho \rfloor + 1}^{\lfloor n\rho \rfloor \vee k} M''_i \right| \\ &\quad + \left| \max_{1 \leq k \leq n} \left[\sum_{i=1}^{\lfloor n\rho \rfloor \wedge k} [(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \mathbb{E}(\mathbf{X}_{i-1}) + (\mu' - \widetilde{\mu})] \right. \right. \\ &\quad \left. \left. + \sum_{i=\lfloor n\rho \rfloor + 1}^{\lfloor n\rho \rfloor \vee k} [(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \mathbb{E}(\mathbf{X}_{i-1}) + (\mu'' - \widetilde{\mu}) - n\psi] \right] \right| \\ &\quad + \max_{1 \leq k \leq n} \left| \sum_{i=1}^k (\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}_n)^\top (\mathbf{X}_{i-1} - \mathbb{E}(\mathbf{X}_{i-1})) \right| \\ &\quad + \max_{1 \leq k \leq n} \left| \sum_{i=1}^k [(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}_n)^\top \mathbb{E}(\mathbf{X}_{k-1}) + (\widetilde{\mu} - \widehat{\mu}_n)] \right|. \end{aligned} \quad (\text{B.5})$$

The second term in (B.5) is deterministic. For its more detailed analysis we need the following lemma.

Lemma B.3. *Under the assumptions of Theorem 4.1 and using the notations from there we have*

$$(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}})^\top \begin{bmatrix} \mathbb{E}(\widetilde{\mathbf{X}}') \\ 1 \end{bmatrix} = \frac{\psi}{\rho} > 0, \quad (\boldsymbol{\theta}'' - \widetilde{\boldsymbol{\theta}})^\top \begin{bmatrix} \mathbb{E}(\widetilde{\mathbf{X}}'') \\ 1 \end{bmatrix} = -\frac{\psi}{1-\rho} < 0$$

with

$$\boldsymbol{\theta}' := \begin{bmatrix} \boldsymbol{\alpha} \\ \mu' \end{bmatrix}, \quad \boldsymbol{\theta}'' := \begin{bmatrix} \boldsymbol{\alpha} \\ \mu'' \end{bmatrix}.$$

Proof. We consider

$$\begin{bmatrix} \boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}} \\ \mu' - \tilde{\mu} \end{bmatrix} = (1 - \rho) \tilde{\mathbf{Q}}^{-1} \mathbf{C}'' \begin{bmatrix} \mathbf{0} \\ \mu' - \mu'' \end{bmatrix}.$$

We can write

$$\begin{bmatrix} \mathbb{E}(\tilde{\mathbf{X}}') \\ 1 \end{bmatrix} = \mathbf{C}' \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix},$$

whence the first equality in the statement is immediate. The second equality can be proven by the same reasoning. Now we need to show that $\psi > 0$. Indeed,

$$\frac{\psi}{\rho} = (1 - \rho)(\mu' - \mu'') \mathbf{e}_{p+1}^\top \mathbf{C}' \tilde{\mathbf{Q}}^{-1} \mathbf{C}'' \mathbf{e}_{p+1},$$

and \mathbf{C}' , \mathbf{C}'' and $\tilde{\mathbf{Q}}^{-1}$ are positive definite matrices. The first two are covariance matrices and because at least one of the offspring and innovation distributions is nondegenerate, they will be positive definite (see C.1). The matrix $\tilde{\mathbf{Q}}^{-1}$ is then the inverse of a convex combination of positive definite matrices, hence positive definite itself. \square

Now the second term in (B.5) can be rewritten as

$$\begin{aligned} & \left| \max_{1 \leq k \leq n} \left[(\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}})^\top \sum_{i=1}^{\lfloor n\rho \rfloor \wedge k} \left[\left(\mathbb{E} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} - \mathbb{E} \begin{bmatrix} \tilde{\mathbf{X}}' \\ 1 \end{bmatrix} \right) \right] + \frac{(\lfloor n\rho \rfloor \wedge k)\psi}{\rho} \right. \right. \\ & \quad \left. \left. + (\boldsymbol{\theta}'' - \tilde{\boldsymbol{\theta}})^\top \sum_{i=\lfloor n\rho \rfloor + 1}^{\lfloor n\rho \rfloor \vee k} \left[\left(\mathbb{E} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} - \mathbb{E} \begin{bmatrix} \tilde{\mathbf{X}}'' \\ 1 \end{bmatrix} \right) \right] - \frac{(k - \lfloor n\rho \rfloor)^+ \psi}{1 - \rho} - n\psi \right] \right| \\ & \leq \max_{1 \leq k \leq n} \left| (\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}})^\top \sum_{i=1}^{\lfloor n\rho \rfloor \wedge k} \left[\left(\mathbb{E} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} - \mathbb{E} \begin{bmatrix} \tilde{\mathbf{X}}' \\ 1 \end{bmatrix} \right) \right] \right| \\ & \quad + \max_{1 \leq k \leq n} \left| (\boldsymbol{\theta}'' - \tilde{\boldsymbol{\theta}})^\top \sum_{i=\lfloor n\rho \rfloor + 1}^{\lfloor n\rho \rfloor \vee k} \left[\left(\mathbb{E} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} - \mathbb{E} \begin{bmatrix} \tilde{\mathbf{X}}'' \\ 1 \end{bmatrix} \right) \right] \right| \\ & \quad + \left| \max_{1 \leq k \leq \lfloor n\rho \rfloor} \left(\frac{(\lfloor n\rho \rfloor \wedge k)}{\rho} - \frac{(k - \lfloor n\rho \rfloor)^+}{1 - \rho} - n \right) \psi \right|. \end{aligned}$$

Because of Lemma A.1, the first two maxima are bounded, and the third one is obviously attained at $k = \lfloor n\rho \rfloor$

with value $\left(\frac{\lfloor n\rho \rfloor}{\rho} - n\right)\psi$, which is also bounded. Therefore the second term in (B.5) is $O(1)$ as $n \rightarrow \infty$. We can use Lemma B.2 to show that the fourth term in (B.5) is $O_P(\sqrt{n})$, but for the first and third terms we also need another result. This can be found as Theorem 3.1. in Kokoszka and Leipus (1998).

Lemma B.4. *Let $(Y_n)_{n \in \mathbb{N}}$ be a sequence of random variables with finite second moments, and let $(c_n)_{n \in \mathbb{N}}$ be a sequence of nonnegative constants. Then, for any $a > 0$,*

$$\begin{aligned} a^2 \mathbb{P} \left(\max_{1 \leq k \leq n} c_k \left| \sum_{j=1}^k Y_j \right| > a \right) &\leq \sum_{k=1}^{n-1} |c_{k+1}^2 - c_k^2| \sum_{i,j=1}^k \mathbb{E}(Y_i Y_j) \\ &\quad + 2 \sum_{k=1}^{n-1} c_{k+1}^2 \left(\mathbb{E}(Y_{k+1}^2) \sum_{i,j=1}^k \mathbb{E}(Y_i Y_j) \right)^{1/2} \\ &\quad + 2 \sum_{k=0}^{n-1} c_{k+1}^2 \mathbb{E}(Y_{k+1}^2). \end{aligned}$$

Lemma B.4 can be applied to the fourth term in the following way to show that it is $O_P(n^{1-\gamma})$:

Lemma B.5. *For a time-homogeneous INAR(p) process satisfying condition \mathbf{C}_0 and any $\gamma < \frac{1}{4}$ we have*

$$\max_{1 \leq k \leq n} k^{\gamma-1} \left\| \sum_{i=1}^k (\mathbf{X}_{k-1} - \mathbb{E}(\mathbf{X}_{k-1})) \right\| = O_P(1).$$

Proof. We will follow the proof of Lemma 4.2 in Hušková et al. (2007) and apply Lemma B.4 with $c_k = k^{\gamma-1}$ and $Y_{i,q} = X_{i-1-q} - \mathbb{E}(X_{i-1-q})$ for $0 \leq q \leq p-1$ to show that the result holds for each component of the vectors. This implies convergence of the 1-norm, and because of the equivalence of vector norms, it is sufficient for the proof of the statement. We have

$$\left| \frac{1}{(k+1)^{2-2\gamma}} - \frac{1}{k^{2-2\gamma}} \right| \leq \frac{2(1-\gamma)}{k^{3-2\gamma}}$$

and

$$\sum_{i=1}^k \sum_{j=1}^k \mathbb{E}(Y_{i,q} Y_{j,q}) = \sum_{i=-q}^{k-1-q} \sum_{j=-q}^{k-1-q} \text{Cov}(X_i, X_j) \leq \kappa k$$

for some constant κ according to Lemma A.2. Therefore,

$$\begin{aligned}
& \sum_{k=1}^{n-1} \left| \frac{1}{(k+1)^{2-2\gamma}} - \frac{1}{k^{2-2\gamma}} \right| \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}(Y_{i,q} Y_{j,q}) + 2 \sum_{k=1}^{n-1} k^{2\gamma-2} \mathbb{E}^{1/2}(Y_{k+1,q}^2) \left(\sum_{i,j=1}^k \mathbb{E}(Y_{i,q} Y_{j,q}) \right)^{1/2} \\
& \quad + 2 \sum_{k=0}^{n-1} k^{2\gamma-2} \mathbb{E}(Y_{k+1,q}^2) \\
& \leq (2\kappa - 2\kappa\gamma) \sum_{k=1}^{n-1} k^{2\gamma-2} + 2(\kappa U_1)^{1/2} \sum_{k=1}^{n-1} k^{2\gamma-3/2} + 2U_1 \sum_{k=0}^{n-1} k^{2\gamma-2},
\end{aligned}$$

where U_1 is the upper boundary of $(\text{Var}(X_n))_{n \in \mathbb{N}}$. The limit of the right hand side as $n \rightarrow \infty$ is finite, which completes the proof. We note the necessity of $\gamma < \frac{1}{4}$ —otherwise the second term in the last expression would not be bounded. This indicates that if we would like to extend Theorem 4.1 to $\gamma = \frac{1}{2}$ we need sharper estimates in place of Lemma B.4. \square

For the estimation of the first term we will use Lemma B.4 again:

Lemma B.6. *For a time-homogeneous INAR(p) process satisfying condition \mathbf{C}_0 and any $\gamma \in (0, \frac{1}{4})$ we have*

$$\max_{1 \leq k \leq n} k^{\gamma-1} \left| \sum_{i=1}^k M_i \right| = O_P(1).$$

Proof. We apply B.4 in the same way as in the proof of Lemma B.5 with $c_k = k^{\gamma-1}$ and $Y_i = M_i$. We note that the M_k are martingale differences, therefore any product $M_i M_j, i \neq j$ has zero mean. Furthermore, the sequence $(\text{Var } M_k)_{k \in \mathbb{N}}$ is clearly bounded, and denoting its upper bound by U , we have

$$\begin{aligned}
& \sum_{k=1}^{n-1} \left| \frac{1}{(k+1)^{2-2\gamma}} - \frac{1}{k^{2-2\gamma}} \right| \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}(Y_i Y_j) + 2 \sum_{k=1}^{n-1} k^{2\gamma-2} \mathbb{E}^{1/2}(Y_{k+1}^2) \left(\sum_{i,j=1}^k \mathbb{E}(Y_i Y_j) \right)^{1/2} \\
& \quad + 2 \sum_{k=0}^{n-1} k^{2\gamma-2} \mathbb{E}(Y_{k+1}^2) \\
& \leq U(2 - 2\gamma) \sum_{k=1}^{n-1} k^{2\gamma-2} + 2U \sum_{k=1}^{n-1} k^{2\gamma-3/2} + 2U \sum_{k=0}^{n-1} k^{2\gamma-2},
\end{aligned}$$

whence the final steps of the proof are the same as in Lemma B.5. \square

Summarizing our results for the terms of (B.5): the second term is $O(1)$, the fourth one is $O_P(n^{1/2})$ and the first and third terms are $O_P(n^{1-\gamma})$, which completes our proof. \square

B.4 Proof of Theorem 4.2

The statement can be written in the form

$$\lim_{K \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P}(|\widehat{\tau}_n - \lfloor n\rho \rfloor| \geq K) = 0,$$

which is equivalent to

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\widehat{\tau}_n - \lfloor n\rho \rfloor| \geq K) = 0.$$

Hence to prove the statement it is enough to show that

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\max_{\lfloor n\rho \rfloor - K < k < \lfloor n\rho \rfloor + K} \sum_{j=1}^k \widehat{M}_j^{(n)} \leq \max_{1 \leq k \leq \lfloor n\rho \rfloor - K} \sum_{j=1}^k \widehat{M}_j^{(n)} \right) = 0, \quad (\text{B.6})$$

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\max_{\lfloor n\rho \rfloor - K < k < \lfloor n\rho \rfloor + K} \sum_{j=1}^k \widehat{M}_j^{(n)} \leq \max_{\lfloor n\rho \rfloor + K \leq k \leq n} \sum_{j=1}^k \widehat{M}_j^{(n)} \right) = 0. \quad (\text{B.7})$$

For (B.6) we consider with a constant K , $K < \lfloor n\rho \rfloor$, the estimate

$$\begin{aligned} & \mathbb{P} \left(\max_{\lfloor n\rho \rfloor - K < k < \lfloor n\rho \rfloor + K} \sum_{j=1}^k \widehat{M}_j^{(n)} \leq \max_{1 \leq k \leq \lfloor n\rho \rfloor - K} \sum_{j=1}^k \widehat{M}_j^{(n)} \right) \\ & \leq \mathbb{P} \left(\sum_{j=1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \leq \max_{1 \leq k \leq \lfloor n\rho \rfloor - K} \sum_{j=1}^k \widehat{M}_j^{(n)} \right) = \mathbb{P} \left(\min_{1 \leq k \leq \lfloor n\rho \rfloor - K} \sum_{j=k+1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \leq 0 \right) \\ & = \mathbb{P} \left(\min_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \leq 0 \right) = \mathbb{P} \left(\min_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \leq 0 \right). \end{aligned}$$

We will use (B.4) again, note that the dominant term is the second one, and estimate the probability with this in mind. For any $K \leq \ell \leq \lfloor n\rho \rfloor$ the expression

$$\ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \quad (\text{B.8})$$

can be decomposed according to (B.4). Now, (B.8) can only be negative in two cases: either the second term in the decomposition is less or equal to $\frac{\psi}{2}$, or it is greater—in which case one of the other three terms has to be less than $-\frac{\psi}{6}$ (for the definition of ψ , see Theorem 4.1).

Now it is clear that, after applying (B.4) to (B.8) we have

$$\begin{aligned}
& \mathbb{P} \left(\min_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} \widehat{M}_j^{(n)} \leq 0 \right) \\
& \leq \mathbb{P} \left(\min_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} [(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})^\top \mathbf{E}(\mathbf{X}_{k-1}) + (\mu' - \widetilde{\mu})] \leq \frac{\psi}{2} \right) \\
& + \mathbb{P} \left(\max_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \left| \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} M_k \right| \geq \frac{\psi}{6} \right) \\
& + \mathbb{P} \left(\max_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \left| \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} (\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}_n)^\top (\mathbf{X}_{k-1} - \mathbf{E}(\mathbf{X}_{k-1})) \right| \geq \frac{\psi}{6} \right) \\
& + \mathbb{P} \left(\max_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \left| \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} [(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}_n)^\top \mathbf{E}(\mathbf{X}_{k-1}) + (\widetilde{\mu} - \widehat{\mu}_n)] \right| \geq \frac{\psi}{6} \right).
\end{aligned}$$

As a consequence of Lemma B.3 and (A.6) the first term can be shown to converge to zero for any K as $n \rightarrow \infty$ with the help of the following simple lemma:

Lemma B.7. *Let $a_n \rightarrow a > 0$, $n \rightarrow \infty$ and $a_i > 0$ for all $i \in \mathbb{N}$. Then*

$$\min_{1 \leq k \leq n} k^{-1} \sum_{i=n-k+1}^n a_i \rightarrow a, \quad n \rightarrow \infty.$$

Proof. First we note that for any $\varepsilon > 0$ and sufficiently large n , we have $\min_{1 \leq k \leq n} k^{-1} \sum_{i=n-k+1}^n a_i < a + \varepsilon$. This can be seen by choosing $k = 1$ for every n . Now we show $\min_{1 \leq k \leq n} k^{-1} \sum_{i=n-k+1}^n a_i > a - \varepsilon$. Let $\nu(\varepsilon)$ be the threshold index so that for $n > \nu(\varepsilon)$ we have $|a_n - a| < \frac{\varepsilon}{2}$. Let us denote by K the sum $\sum_{i=1}^{\nu(\varepsilon)} a_i$. Clearly,

$$\left| \min_{1 \leq k \leq n - \nu(\varepsilon)} k^{-1} \sum_{i=n-k+1}^n a_i - a \right| < \frac{\varepsilon}{2}.$$

Furthermore, for any $n > k > n - \nu(\varepsilon)$ we have

$$k^{-1} \sum_{i=n-k+1}^n a_i > n^{-1} \sum_{i=\nu(\varepsilon)}^n a_i = \frac{n - \nu(\varepsilon)}{n} \left[(n - \nu(\varepsilon))^{-1} \sum_{i=\nu(\varepsilon)}^n a_i \right].$$

For sufficiently large n the first factor is close to 1, and the second factor is closer to a than $\frac{\varepsilon}{2}$ for every n . This suffices for the proof. \square

Because of (A.6) and Lemma B.2, the fourth term also converges to zero for all K as $n \rightarrow \infty$. Indeed,

$(\tilde{\alpha} - \hat{\alpha}_n) \xrightarrow{P} 0$ and

$$\max_{K \leq \ell \leq \lfloor n\rho \rfloor} \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} E(\mathbf{X}_{k-1}) \leq \max_{1 \leq k \leq \lfloor n\rho \rfloor} E(\mathbf{X}_k)$$

and due to (A.6) the right hand side is bounded as $n \rightarrow \infty$. The same reasoning applies to $(\tilde{\mu} - \hat{\mu}_n)$. The convergence of the second and third terms is summarized in the following lemma.

Lemma B.8. *For a time-homogeneous INAR(p) process satisfying condition \mathbf{C}_0 we have for any $a > 0$,*

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \left| \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} (\mathbf{X}_{j-1} - E(\mathbf{X}_{j-1})) \right| > a \right) = 0$$

and

$$\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{K \leq \ell \leq \lfloor n\rho \rfloor - 1} \left| \ell^{-1} \sum_{j=\lfloor n\rho \rfloor - \ell + 1}^{\lfloor n\rho \rfloor} M_j \right| > a \right) = 0.$$

Proof. Similarly to the proof of Lemma B.5 we will again employ Lemma B.4 with $c_k = (K + k - 1)^{-1}$ and $Y_{1,q} = \sum_{j=\lfloor n\rho \rfloor - K + 1}^{\lfloor n\rho \rfloor} X_{j-q}$ and $Y_{i,q} = X_{\lfloor n\rho \rfloor - K + 1 - i - q}$ for $i \geq 2$ and $0 \leq q \leq p - 1$.

By an easy calculation

$$\sum_{i,j=1}^k E(Y_i Y_j) = \sum_{i,j=\lfloor n\rho \rfloor - K - k + 1}^{\lfloor n\rho \rfloor} E((\mathbf{X}_{i-1} - E(\mathbf{X}_{i-1}))(\mathbf{X}_{j-1} - E(\mathbf{X}_{j-1}))).$$

Therefore, applying the same estimations and notations as in the proof of Lemma B.5 with $\gamma = 0$, we obtain the following upper limit for the probability in question:

$$2\kappa \sum_{\ell=K}^{\lfloor n\rho \rfloor - 1} (\ell + 1)^{-2} + U_1 \sum_{\ell=K}^{\lfloor n\rho \rfloor - 1} (\ell + 1)^{-3/2} + \frac{U_1}{K} + U_1 \sum_{\ell=K-1}^{\lfloor n\rho \rfloor - 1} (\ell + 1)^{-2}.$$

It is obvious that as $n \rightarrow \infty$ and then $K \rightarrow \infty$, the above expression converges to 0, which suffices for our proof. For the second statement the arguments are the same. We note that $(M_n)_{n \in \mathbb{N}}$ is a martingale difference sequence, hence its elements are pairwise uncorrelated. Furthermore, $\text{Var}(M_n)_{n \in \mathbb{N}}$ is bounded, which implies $\text{Var}(M_1 + \dots + M_n) = O(n)$ immediately. \square

To prove (B.7) the proof is analogous with one exception: in place of Lemma B.7 we need the following result.

Lemma B.9. *Let $a_n \rightarrow a > 0$, $n \rightarrow \infty$ and $a_i > 0$ for all $i \in \mathbb{N}$. Then*

$$\lim_{K \rightarrow \infty} \inf_{k \geq K} k^{-1} \sum_{i=1}^k a_i = a, \quad n \rightarrow \infty.$$

Proof. We only need to observe that convergence of a_n implies convergence in Cesaro mean as well, therefore, for a sufficiently large K and for all $k \geq K$ the average $k^{-1} \sum_{i=1}^k a_i$ is close to a . \square

B.5 Adapting the proofs of Theorem 4.1 and Theorem 4.2 to other forms of the alternative hypothesis

The proofs of Theorem 4.1 and Theorem 4.2 heavily exploit the relatively simple structure of the test process for detecting change in μ only. For the other parameters, even the limit in Theorem 4.1 will be different, namely, we have the following theorem.

Theorem B.10. *Suppose that H_A holds with $\tau = \max(\lfloor n\rho \rfloor, 1)$, $\rho \in (0, 1)$, and the change is only in α_q , namely, α_q changes from α'_q to α''_q , where $\alpha' > \alpha'' > 0$. Suppose, furthermore, that condition C_A holds. Then for any $\gamma \in (0, \frac{1}{4})$ we have*

$$\max_{1 \leq k \leq n} \sum_{j=1}^k \widehat{M}_j^{(n)} X_{j-q} = n\psi_q + O_P(n^{1-\gamma}) \quad \text{as } n \rightarrow \infty,$$

with

$$\psi_q := \rho(1 - \rho)(\alpha'_q - \alpha''_q) \mathbf{e}_q^\top \mathbf{C}'' \tilde{\mathbf{Q}}^{-1} \mathbf{C}' \mathbf{e}_q > 0,$$

where \mathbf{e}_q is the q -th $(p+1)$ -dimensional unit vector.

The proof is analogous to that of Theorem 4.1. All the techniques in the proofs can be directly adapted, noting that all means converge due to (A.6), and Lemma B.2 will remain true under any form of the alternative hypothesis. In place of (B.4), we can then write

$$\begin{aligned} \widehat{M}_k^{(n)} X_{k-q} &= M'_k X_{k-q} + [(\boldsymbol{\alpha}' - \tilde{\boldsymbol{\alpha}})^\top \mathbf{E}(\mathbf{X}_{k-1} X_{k-q}) + (\mu' - \tilde{\mu}) \mathbf{E}(X_{k-q})] \\ &\quad + (\boldsymbol{\alpha}' - \hat{\boldsymbol{\alpha}}_n)^\top (\mathbf{X}_{k-1} X_{k-q} - \mathbf{E}(\mathbf{X}_{k-1} X_{k-q})) + (\mu' - \tilde{\mu}) (X_{k-q} - \mathbf{E}(X_{k-q})) \\ &\quad + [(\tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_n)^\top \mathbf{E}(\mathbf{X}_{k-1} X_{k-q}) + (\tilde{\mu} - \hat{\mu}_n) \mathbf{E}(X_{k-q})]. \end{aligned} \tag{B.9}$$

As in (B.4), the second term will be dominant here and in place of Lemma B.3 we have

$$(\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}})^\top \begin{bmatrix} \mathbf{E}(\tilde{\mathbf{X}}' \tilde{X}'_{-q+1}) \\ \mathbf{E}(\tilde{X}'_{-q+1}) \end{bmatrix} = \frac{\psi_q}{\rho} > 0, \quad (\boldsymbol{\theta}'' - \tilde{\boldsymbol{\theta}})^\top \begin{bmatrix} \mathbf{E}(\tilde{\mathbf{X}}'' \tilde{X}''_{-q+1}) \\ \mathbf{E}(\tilde{X}''_{-q+1}) \end{bmatrix} = -\frac{\psi_q}{1 - \rho},$$

We need (ii) in Lemma A.2 to be able to apply Lemma B.4 to the first and third terms in (B.9). The rest of the adaptation is straightforward.

If there is a change in multiple parameters, the analogue of Theorem 4.1 can still be stated, but the sign of the dominant term depends nontrivially on the directions of the changes in these parameters (i.e., in Lemma B.3 we can determine the limit but we cannot determine its sign without calculating it explicitly). The one-sided test, however, relies on our knowledge of the sign of the dominant term. This warns us that we should use the one-sided tests only when testing for change in a single parameter.

C Technical details

C.1 Invertibility of the matrices \mathbf{Q}_n , \mathbf{C}' and \mathbf{C}''

In (2.2) we assumed that the matrix \mathbf{Q}_n is invertible, and similarly, in B.2 we assumed that \mathbf{C}' and \mathbf{C}'' are positive definite. The following two lemmas will show that these assumptions are correct.

Lemma C.1. *For a homogeneous INAR(p) process with $\mu > 0$, for which either $\alpha_q \in (0, 1)$ for some $q \in \{1, 2, \dots, p\}$ or $\sigma > 0$, we have*

$$\mathbb{P}(\mathbf{Q}_n \text{ is singular}) \rightarrow 0.$$

Proof. Since

$$\mathbf{Q}_n = \sum_{i=1}^n \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix}^\top$$

is a sum of positive semidefinite matrices, it is positive semidefinite itself. Therefore, its singularity is equivalent to the condition that for some $0 \neq \mathbf{v} \in \mathbb{R}^{p+1}$ and every index $i \in \{1, \dots, n\}$, we have

$$\mathbf{v}^\top \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix}^\top \mathbf{v} = 0,$$

which is equivalent to the condition that the linear span of

$$\left\{ \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix}, i = 1, \dots, n \right\}$$

is a proper subspace of \mathbb{R}^{p+1} . Now, using the continuity of probability, our statement is equivalent to the following:

$$\mathbb{P} \left(\text{span} \left\{ \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix}, i \in \mathbb{N} \right\} < \mathbb{R}^{p+1} \right) = 0, \tag{C.1}$$

where $<$ denotes proper subspace. For simplicity, throughout the proof we will use the notation

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{X}_{i-1} \\ 1 \end{bmatrix}, \quad i \in \mathbb{N}.$$

It is clear that all values of $(\mathbf{Y}_i)_{i \in \mathbb{N}}$ fall into $\mathbb{N}_0^p \times \{1\}$. We introduce the following notation for the set of spaces that can be spanned by the values of the process:

$$\mathcal{S} := \{S < \mathbb{R}^{p+1} : S = \text{span}\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n \in \mathbb{N}_0^p \times \{1\}, n \in \mathbb{N}\}. \quad (\text{C.2})$$

We can notice that \mathcal{S} is countable. Indeed, every generating system of a subspace contains a basis, therefore every subspace $S \in \mathcal{S}$ has a basis whose elements are from $\mathbb{N}_0^p \times \{1\}$. Such a basis is from $(\mathbb{N}_0^p \times \{1\})^k$ where $k = \dim S$, and $0 \leq k \leq p$, and of course a basis corresponds to only one subspace. Now, since $\mathbb{N}_0^p \times \{1\}$ is countable, $(\mathbb{N}_0^p \times \{1\})^k$ is also countable for any $k \in \mathbb{N}$, therefore $\cup_{k=0}^p (\mathbb{N}_0^p \times \{1\})^k$ is also countable, and so is \mathcal{S} .

Now we reformulate the event in (C.1):

$$\begin{aligned} \{\text{span}\{\mathbf{Y}_i, i \in \mathbb{N}\} < \mathbb{R}^{p+1}\} &= \bigcup_{S < \mathbb{R}^{p+1}} \{\text{span}\{\mathbf{Y}_i, i \in \mathbb{N}\} = S\} = \bigcup_{S \in \mathcal{S}} \{\text{span}\{\mathbf{Y}_i, i \in \mathbb{N}\} = S\} \\ &\subseteq \bigcup_{S \in \mathcal{S}} \{\text{span}\{\mathbf{Y}_i, i \in \mathbb{N}\} \subseteq S\}. \end{aligned} \quad (\text{C.3})$$

Since the last union is countable, we can apply σ -subadditivity to show (C.1) if we can prove

$$\mathbb{P}(\text{span}\{\mathbf{Y}_i, i \in \mathbb{N}\} \subseteq S) = \lim_{n \rightarrow \infty} \mathbb{P}(\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\} \subseteq S) = 0, \quad \forall S \in \mathcal{S}. \quad (\text{C.4})$$

Here the first equality is trivial by the continuity of probability; it is the second equality which requires a more detailed proof. The first step in the proof of (C.4) relies on the mechanism by which the components of \mathbf{Y}_{i+1} can be obtained from those of \mathbf{Y}_i . For a fixed $S \in \mathcal{S}$ the elements of \mathcal{S} can be viewed as the solutions of a homogeneous system of independent linear equations, i.e., $\mathbf{y} \in S$ if and only if

$$\sum_{j=1}^{p+1} \lambda_{i,j} y^{(j)} = 0, \quad i = 1, 2, \dots, p+1 - \dim S. \quad (\text{C.5})$$

This representation is not unique, but we can fix one such representation. Now let us introduce

$$K(S) := \min\{j \in \{1, 2, \dots, p\} : \max_i |\lambda_{i,j}| > 0\} \text{ and}$$

$$c(s) := \min\{i \in \{1, 2, \dots, p+1 - \dim S\} : |\lambda_{i,K(S)}| > 0\}.$$

This notation means that $K(S)$ is the first column index for which a nonzero coefficient appears in some equation in (C.5) and the first nonzero coefficient in the $c(s)$ -th equation has index $K(S)$. We also note that $K(S) = p+1$ is impossible because that would mean that the only equation is $y^{(p+1)} = 0$, which does not hold for any element of $\mathbb{N}_0^p \times \{1\}$.

Let us now fix an arbitrary $i \in \mathbb{N}$ and $\omega \in \Omega$ from our underlying probability space such that $\mathbf{Y}_i(\omega) = \mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(p)}, 1)^\top$. Then we have

$$\mathbf{Y}_{i+K(S)}(\omega) = \left(X_{i+K(S)-1}(\omega), \dots, X_i(\omega), y^{(1)}, \dots, y^{(p-K(S))}, 1 \right)^\top \quad (\text{see (A.1)}).$$

Hence, for $\mathbf{Y}_{i+K(S)}(\omega) \in S$ to hold, it is necessary (but usually not sufficient) that $\mathbf{Y}_{i+K(S)}(\omega)$ satisfy the $c(s)$ -th equation in (C.5), i.e.,

$$\begin{aligned} \sum_{j=1}^{K(S)} \lambda_{c(s),j} X_{i+K(S)-j}(\omega) + \sum_{j=K(S)+1}^p \lambda_{c(s),j} y^{(j-K(S))} + \lambda_{c(s),p+1} &= 0 \\ \Leftrightarrow \lambda_{c(s),K(S)} X_i(\omega) + \sum_{j=K(S)+1}^p \lambda_{c(s),j} y^{(j-K(S))} + \lambda_{c(s),p+1} &= 0. \end{aligned}$$

This linear equation has a unique solution for $X_i(\omega)$ because $\lambda_{c(s),K(S)} \neq 0$. Let us denote this unique solution by $m(\mathbf{y}, S)$ (by simple algebraic considerations one can see that this quantity does not depend on the representation on (C.5), but this is not necessary to our proof). Therefore, if $X_i(\omega) \neq m(\mathbf{y}, S)$, then $\omega \notin \{\mathbf{Y}_{i+K(S)} \in S, \mathbf{Y}_i = \mathbf{y}\}$, hence

$$\{\mathbf{Y}_{i+K(S)} \in S, \mathbf{Y}_i = \mathbf{y}\} \subseteq \{X_i = m(\mathbf{y}, S), \mathbf{Y}_i = \mathbf{y}\} \quad \forall i \in \mathbb{N}, \quad \forall \mathbf{y} \in \mathbb{N}_0^p \times \{1\}.$$

If $m(\mathbf{y}, S) \notin \mathbb{N}$, then we have $\{\mathbf{Y}_{i+K(S)} \in S, \mathbf{Y}_i = \mathbf{y}\} = \emptyset$.

Now we will consider the second event in (C.4) for $n = n + K(S)$, and split it according to the initial value of the process:

$$\{\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n+K(S)}\} \subseteq S\} = \bigcup_{\mathbf{y}_1 \in \mathbb{N}_0^p \times \{1\}} \{\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n+K(S)}\} \subseteq S, \mathbf{Y}_1 = \mathbf{y}_1\}. \quad (\text{C.6})$$

The individual events in the union can be transformed in the following way:

$$\begin{aligned}
\{\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n+K(S)}\} \subseteq S, \mathbf{Y}_1 = \mathbf{y}_1\} &= \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{n+K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1\} \\
&= \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_{1+K(S)} \in S, \dots, \mathbf{Y}_{n+K(S)} \in S\} \\
&\subseteq \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, X_1 = m(\mathbf{y}_1, S), \mathbf{Y}_{2+K(S)} \in S, \dots, \mathbf{Y}_{n+K(S)} \in S\} \\
&= \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_{2+K(S)} \in S, \dots, \mathbf{Y}_{n+K(S)} \in S\} \\
&\subseteq \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, X_2 = m(\mathbf{y}_2, S), \mathbf{Y}_{3+K(S)} \in S, \dots, \mathbf{Y}_{n+K(S)} \in S\} \\
&= \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_3 = \mathbf{y}_3, \mathbf{Y}_{3+K(S)} \in S, \dots, \mathbf{Y}_{n+K(S)} \in S\} \\
&\vdots \\
&\subseteq \{\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S, \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_3 = \mathbf{y}_3, \dots, \mathbf{Y}_n = \mathbf{y}_n\},
\end{aligned} \tag{C.7}$$

where the sequence $(\mathbf{y}_i)_{i=1}^n$ is defined by the recursion

$$\mathbf{y}_i = \begin{bmatrix} m(\mathbf{y}_{i-1}, S) \\ \mathbf{y}_{i-1}^{(1)} \\ \vdots \\ \mathbf{y}_{i-1}^{(p-1)} \\ 1 \end{bmatrix}, \quad i = 2, 3, \dots, n. \tag{C.8}$$

We would like to represent the probability of the last event in (C.7) as a product of transition probabilities. For this we first need to determine whether the event is empty, and now we will give the necessary conditions on \mathbf{y}_1 for its nonemptiness. The first condition is, clearly, that all elements of the sequence defined in (C.8) fall into $\mathbb{N}_0^p \times \{1\}$. We will not investigate this condition in any further detail, we only note that this imposes a deterministic condition on \mathbf{y}_1 . Another deterministic condition is that $\mathbf{Y}_1 \in S, \mathbf{Y}_2 \in S, \dots, \mathbf{Y}_{K(S)} \in S$ should all hold. Because the first $K(S) - 1$ coefficients are all zero in any equation in (C.5) and \mathbf{Y}_1 contains all the components indexed $K(S)$ or greater in $\mathbf{Y}_1, \dots, \mathbf{Y}_{K(S)}$, the validity of these inclusions is determined by \mathbf{y}_1 alone. This imposes the second (again, deterministic) condition on \mathbf{y}_1 . If we denote the set of \mathbf{y}_1 which fulfill both these conditions by U_n , we have from (C.6) and (C.7),

$$\{\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n+K(S)}\} \subseteq S\} \subseteq \bigcup_{\mathbf{y}_1 \in U_n} \{\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_3 = \mathbf{y}_3, \dots, \mathbf{Y}_n = \mathbf{y}_n\},$$

hence by σ -subadditivity (U_n is clearly countable),

$$\begin{aligned} \mathbb{P}(\text{span}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n+K(S)}\} \subseteq S) &\leq \sum_{\mathbf{y}_1 \in U_n} \mathbb{P}(\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2, \dots, \mathbf{Y}_n = \mathbf{y}_n) \\ &= \sum_{\mathbf{y}_1 \in U_n} \mathbb{P}(\mathbf{Y}_1 = \mathbf{y}_1) p_{\mathbf{y}_1, \mathbf{y}_2} p_{\mathbf{y}_2, \mathbf{y}_3} \cdots p_{\mathbf{y}_{n-1}, \mathbf{y}_n}, \end{aligned} \quad (\text{C.9})$$

where $p_{\mathbf{u}, \mathbf{v}}$ denotes the transition probability of \mathbf{Y} from \mathbf{u} to \mathbf{v} . Because the sets $(U_n)_{n \in \mathbb{N}}$ form a nonincreasing sequence (the second condition does not depend on n , and the first one become more restrictive as n increases), it is sufficient to show that for any sequence $(\mathbf{y}_i)_{i \in \mathbb{N}} \in (\mathbb{N}_0^p \times \{1\})^{\mathbb{N}}$ we have

$$\lim_{n \rightarrow \infty} p_{\mathbf{y}_1, \mathbf{y}_2} p_{\mathbf{y}_2, \mathbf{y}_3} \cdots p_{\mathbf{y}_{n-1}, \mathbf{y}_n} = 0, \quad (\text{C.10})$$

and from this we will get (C.4). For the proof of (C.10) we will need to establish upper bounds for the transition probabilities. We will first consider the case when $\sigma > 0$, i.e., when the innovation distribution is nondegenerate.

Let us fix $\mathbf{u}, \mathbf{v} \in \mathbb{N}_0^p \times \{1\}$ so that $v^{(2)} = u^{(1)}, v^{(3)} = u^{(2)}, \dots, v^{(p)} = u^{(p-1)}, v^{(p+1)} = 1$. We would like to give an upper bound for $p_{\mathbf{u}, \mathbf{v}}$. We have for every $i \in \mathbb{N}$ and any $m \in \mathbb{N}_0$,

$$\begin{aligned} \mathbb{P}\left(\mathbf{Y}_{i+1} = \mathbf{v} \mid \mathbf{Y}_i = \mathbf{u}, \sum_{j=1}^p \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) &= \mathbb{P}\left(\varepsilon_i = v^{(1)} - m \mid \mathbf{Y}_i = \mathbf{u}, \sum_{j=1}^p \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) \\ &= \mathbb{P}\left(\varepsilon_i = v^{(1)} - m\right). \end{aligned}$$

Applying the law of total probability we get

$$p_{\mathbf{u}, \mathbf{v}} = \sum_{m \in \mathbb{N}_0} \mathbb{P}(\varepsilon_i = v^{(1)} - m) \mathbb{P}\left(\sum_{j=1}^p \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) \leq \max_{k \in \mathbb{N}_0} \mathbb{P}(\varepsilon_i = k) < 1, \quad (\text{C.11})$$

since the innovation distribution was nondegenerate. Therefore, if $\sigma > 0$, then we have a uniform upper bound on the transition probabilities, which implies (C.10) immediately.

The other case is if the innovation distribution is degenerate. First we note that in this case the innovation is equal to its expectation $\mu > 0$ almost surely, so that all components of \mathbf{Y}_i are positive for $i \geq p+1$. According to the conditions, there is a coefficient $\alpha_q, q \in \{1, 2, \dots, p\}$ such that $0 < \alpha_q < 1$. Similarly to the previous reasoning, if additionally we suppose that all components of \mathbf{u} and \mathbf{v} are greater or equal to μ , we

have

$$\begin{aligned}
& \mathbb{P} \left(\mathbf{Y}_{i+1} = \mathbf{v} \mid \mathbf{Y}_i = \mathbf{u}, \mu + \left(\sum_{\substack{j=1 \\ j \neq q}}^p \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} \right) + \sum_{\ell=1}^{u^{(q)}-1} \xi_{q,i,\ell} = m \right) \\
&= \mathbb{P} \left(\xi_{q,i,u^{(q)}} = v^{(1)} - m \mid \mathbf{Y}_i = \mathbf{u}, \mu + \left(\sum_{\substack{j=1 \\ j \neq q}}^p \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} \right) + \sum_{\ell=1}^{u^{(q)}-1} \xi_{q,i,\ell} = m \right) \\
&= \mathbb{P} \left(\xi_{q,i,u^{(q)}} = v^{(1)} - m \right).
\end{aligned}$$

Here we note that $\xi_{q,i,u^{(q)}}$ is a meaningful notation because $u^{(q)} \geq \mu$ and μ is a positive integer. Applying the law of total probability again, we have that

$$p_{\mathbf{u},\mathbf{v}} \leq \max(\alpha_q, 1 - \alpha_q) < 1,$$

which again gives a uniform upper bound for the transition probabilities and yields (C.10). With this our proof is complete. \square

It may be worth noting that Lemma C.1 imposes very weak conditions on the process—we only neglect the trivial case when all innovation and offspring distributions are degenerate. Also, the lemma does not require that the process start from zero—the initial distribution can be arbitrarily chosen on U . This gives us a chance to prove two important corollaries.

Corollary C.2. *For an INAR(p) process under the alternative hypothesis satisfying the assumptions of Lemma C.1 both before and after the change, and $\tau = \max(1, \lfloor n\rho \rfloor)$ for some $\rho > 0$ constant, we have*

$$\mathbb{P}(\mathbf{Q}_n \text{ is singular}) \rightarrow 0.$$

Proof. To show this statement we only need to note that due to Lemma C.1 we have

$$\mathbb{P}(\mathbf{Q}_{\lfloor n\rho \rfloor} \text{ is singular}) \rightarrow 0,$$

and clearly

$$\{\mathbf{Q}_n \text{ is singular}\} \subseteq \{\mathbf{Q}_{\lfloor n\rho \rfloor} \text{ is singular}\}$$

due to the reasoning at the beginning of the proof of Lemma C.1. \square

Remark. The conditions of this corollary could be greatly relaxed. Actually, it can be shown to be true for any change point sequence τ_n , but a proper formulation is tedious and not a goal of this paper.

Corollary C.3. *Under the conditions of Theorem 4.1 both \mathbf{C}' and \mathbf{C}'' are positive definite.*

Proof. First we prove for \mathbf{C}' . We note that Lemma C.1 did not impose any conditions on the initial distribution of the process \mathbf{Y} , therefore we can start the process from its stationary distribution (the existence of which is a trivial corollary of the existence of such a distribution for \mathbf{X} before the change). Now, the singularity of \mathbf{C}' is equivalent to the condition

$$\mathbb{P} \left(\begin{bmatrix} \tilde{X}' \\ 1 \end{bmatrix} \in S \right) < 1, \quad \forall S < \mathbb{R}^{p+1}.$$

Let us now suppose that the stationary distribution of \mathbf{Y} is concentrated on a proper subspace $S < \mathbb{R}^{p+1}$. From (C.4), however, we conclude that the probability of the process remaining in S forever is zero. As the distribution of \mathbf{Y}_n is the stationary distribution for every time n , this is an immediate contradiction. Therefore \mathbf{C}' is nonsingular, but since it is a covariance matrix, it is positive semidefinite, therefore it has to be positive definite. The proof is the same for \mathbf{C}'' . \square

C.2 The conditional moments of M_k

We shall now derive several moments of M_k conditionally on \mathcal{F}_{k-1} (this calculation is a reproduction of that in T. Szabó (2011a)). Let us write M_k in the form

$$M_k = \sum_{j=1}^{X_{k-1}} (\xi_{1,k,j} - \alpha_1) + \sum_{j=1}^{X_{k-2}} (\xi_{2,k,j} - \alpha_2) + \dots + \sum_{j=1}^{X_{k-p}} (\xi_{p,k,j} - \alpha_p) + (\varepsilon_k - \mu).$$

All the terms on the right hand side have zero mean and are independent of each other conditionally on \mathcal{F}_{k-1} , therefore

$$\mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) = \alpha_1(1 - \alpha_1)X_{k-1} + \dots + \alpha_p(1 - \alpha_p)X_{k-p} + \sigma^2.$$

Similarly,

$$\begin{aligned}
E(M_k^4 | \mathcal{F}_{k-1}) &= \sum_{i=1}^p E((\xi_{i,1,1} - \alpha_i)^4) X_{k-i} + 3 \sum_{i,j=1, i \neq j}^p E((\xi_{i,1,1} - \alpha_i)^2 (\xi_{j,1,1} - \alpha_j)^2) X_i X_j \\
&\quad + 6 \sum_{i=1}^p \binom{X_i}{2} E^2((\xi_{i,1,1} - \alpha_i)^2) + 6 \sum_{i=1}^p X_{k-i} E((\xi_{i,1,1} - \alpha_i)^2 (\varepsilon_1 - \mu)^2) + E((\varepsilon_1 - \mu)^4) \\
&= \boldsymbol{\alpha}_4^\top X_{k-i} + 3 \sum_{i,j=1, i \neq j}^p \alpha_i (1 - \alpha_i) \alpha_j (1 - \alpha_j) X_i X_j \\
&\quad + 6 \sum_{i=1}^p \binom{X_i}{2} \alpha_i^2 (1 - \alpha_i)^2 + 6 \boldsymbol{\alpha}_2^\top \mathbf{X}_{k-1} \sigma^2 + E((\varepsilon_1 - \mu)^4).
\end{aligned}$$

Acknowledgements

The authors wish to thank Edit Gombay at the University of Alberta and Lajos Horváth at the University of Utah for drawing their attention to this research topic and directing them towards some very useful literature.

We are deeply grateful to several anonymous reviewers for a thorough analysis of the manuscript and insightful comments which greatly improved the paper.

The research was partially supported by the Hungarian Scientific Research Fund under Grant No. OTKA T-079128, the Hungarian–Chinese Intergovernmental S & T Cooperation Programme for 2011-2013 under Grant No. 10-1-2011-0079 and the TÁMOP-4.2.2/B-10/1-2010-0012 project.

References

- Barczy, M., Ispány, M., & Pap, G. (2011). Asymptotic behavior of unstable INAR(p) processes. *Stochastic Process. Appl.*, 121(3), 583–608.
- Barczy, M., Ispány, M., & Pap, G. (2012). *Asymptotic behavior of CLS estimators for unstable INAR(2) models*. (arXiv:1202.1617v1)
- Chung, K. L. (1960). *Markov Chains with Stationary Transition Probabilities*. Berlin: Springer-Verlag.
- Csörgő, M., & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: Wiley.
- Davis, R. A., Dunsmuir, W. T. M., & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3), 491–505.
- Davis, R. A., & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, 96(3), 735–749.

- Du, J.-G., & Li, Y. (1991). The integer valued autoregressive (INAR(p)) model. *J. Time Series Anal.*, 12(2), 129–142.
- Gombay, E. (2008). Change detection in autoregressive time series. *J. Multivariate Anal.*, 99(3), 451–464.
- Hušková, M., Prášková, Z., & Steinebach, J. (2007). On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist. Plann. Inference*, 137(4), 1243–1259.
- Hyndman, R. J. (n.d.). *Time Series Data Library*. Retrieved 16 Aug, 2012, from <http://data.is/TSDLDemo>
- Kang, J., & Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *J. Time Series Anal.*, 30(2), 239–258.
- Klimko, L. A., & Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, 6(3), 629–642.
- Kokoszka, P., & Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40(4), 385–393.
- Kuiper, N. H. (1960). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A*, 63(1), 38–47.
- Meyn, S., & Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability* (2nd ed.). Cambridge: Cambridge University Press.
- Orey, S. (1971). *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities* (2nd ed.). London: Van Nostrand Reinhold Company.
- Quine, M. P. (1970). The multi-type Galton-Watson process with immigration. *J. Appl. Probab.*, 7(2), 411–422.
- Silva, I. (2005). *Contributions to the analysis of discrete-valued time series*. Unpublished doctoral dissertation, Departamento de Matematica Aplicada, Faculdade de Ciencias da Universidade do Porto. Retrieved 20 Aug, 2012, from <http://paginas.fe.up.pt/~ims/tese.htm>
- T. Szabó, T. (2011a). *Change point detection for integer-valued autoregressive (INAR(p)) processes*. (Paper presented at the Hungarian National Conference of Scientific Students' Associations (OTDK))
- T. Szabó, T. (2011b). Test Statistics for Parameter Changes in INAR(p) Models and a Simulation Study. *Austrian Journal of Statistics*, 40(4), 265–280.
- Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton–Watson process with immigration. *Adv. in Appl. Probab.*, 14(1), 1–20.